CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

General Facts

- A random variable X is said to be continuous if its range is an interval $(a, b) \subseteq \mathbb{R}$, or a collection of intervals.
- Continuous r.v.'s don't exist in "real life", however they may represent useful approximations in many cases. Examples include:
  - Measuring occurrences of random events in time
  - Measuring distance or height
  - Determining the angle a spinner point comes to rest at.

Suppose X is a continuous random variable with range $[0, 1]$ (for example). How can we describe probabilities associated with X?

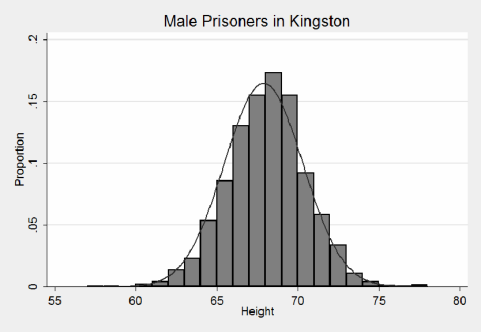
- we can try to define the analog of the p.f. $f_X(x) = P(X = x)$. In order to satisfy the axioms of probability we would need that

$$\sum_{x \in [0,1]} P(X = x) = 1,$$

  but this is hard to make sense of. If $P(X = x) > 0$ for all $x \in [0, 1]$, how can this sum converge?

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- Another idea comes from trying to model relative frequency from the histogram.



$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

Definition. We say that a continuous random variable X has probability density function $f_X(x)$ if

1. $$f_X(x) \geq 0 \text{ for all } x.$$
2. $$\int_{-\infty}^{\infty} f_X(x)dx = 1$$
3. For any a and b (including $a = -\infty$ and $b = \infty$) we have
$$P(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

The probability density function is often abbreviated pdf.

- We can define a continuous r.v. X as a probability model by specifying its pdf.
  This is similar to the case of a discrete random variable, where we specify a model by defining its pf.
- Compare:
  - If X is a discrete r.v. with the pf f_X, then for any event A

$$P(X \in A) = \sum_{x:x \in A} f_X(x).$$

  - If X is a continuous r.v. with the pdf f_X then for any event A[1]

$$P(X \in A) = \int_{x:x \in A} f_X(x)dx \equiv \int_A f_X(x)dx. \qquad (1)$$

[1] Formally this works only for "measurable" events, but we are going to ignore this fact.

Example. Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

(i) Compute c so that this is a valid pdf
(ii) Graph $f(x)$
(iii) Compute $P(X \geq 1/2)$
(iv) Compute $P(1/4 \leq X \leq 3/4)$
(v) Compute $P(X = 1/2)$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Remark. Notice that if X has pdf $f(x)$, then by (1)

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0.$$

In other words, continuous random variables do not take any one value with positive probability.

This implies that for continuous random variables

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b).$$

However, the shape of a pdf matters:

$$
\begin{aligned}
P(X \in (x - \frac{\Delta}{2}, x + \frac{\Delta}{2})) &= \int_{x-\Delta/2}^{x+\Delta/2} f(u)du \\
&\approx f(x)\Delta.
\end{aligned}
$$

Thus, if the pdf of a rv X is relatively large at x, then X will take values more often in a neighborhood of x.

Clicker Question(s).

Cumulative Distribution Function

The distribution of a continuous rv can also be defined by specifying its CDF.

Definition. The CDF of a random variable X is

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

- This is the same definition as for discrete r.v's, but if X is continuous with pdf $f(x)$, then

$$F_X(x) := P(X \in (-\infty, x]) = \int_{-\infty}^{x} f(y)dy \quad x \in \mathbb{R}.$$

  Moreover, by the fundamental theorem of calculus,

$$\frac{d}{dx}F_X(x) \equiv F'_X(x) = f(x).$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- The CDF of a continuous random variable is often easier to calculate, and less difficult to work with, than for a discrete rv.

  Typically we use it to calculate probabilities of the following form:

$$\begin{aligned} P(a \leq X \leq b) &= P(X \in (-\infty, b]) - P(X \in (-\infty, a)) \\ &= \int_{-\infty}^{b} f(y)dy - \int_{-\infty}^{a} f(y)dy \\ &= F_X(b) - F_X(a). \end{aligned}$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Properties

Properties of the CDF of a continuous random variable:

1. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.
2. $F(x)$ is continuous.
3. $F(x)$ is differentiable (possibly except a countable number of points).
4. $F(x)$ is non-decreasing.

   The set of arguments for which a pdf is strictly positive is called “the support of the pdf”:

   $$supp(f) = \{x \in \mathbb{R} : f(x) > 0\}.$$

   From the definition of CDF, it follows that

   $$F(x) \text{ is strictly increasing only for } x \in supp(f).$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Example. Suppose that X is a continuous random variable with the probability density function

$$f_X(x) = \begin{cases} x & \text{if } 0 \leq x \leq \sqrt{2}, \\ 0 & \text{otherwise.} \end{cases}$$

a) Compute F_X
b) Graph F_X
c) Compute $P(1/2 \leq X < 1)$.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Quantiles

- Suppose X is a continuous random variable with the cumulative distribution function $F(x)$. Let $p \in (0,1)$.

  Definition. The pth quantile of X (or the pth quantile of the distribution) is the value $q(p)$, such that

  $$P(X \leq q(p)) = p,$$

  or, in terms of the CDF,

  $$F(q(p)) = p.$$

  - The value $q(p)$ is also called the 100*pth* percentile of the distribution.

- If F is strictly increasing, then

$$q(p) = F^{-1}(p),$$

where F^{-1} is the inverse function of F, that is

$$F^{-1}(F(x)) = x \quad \text{and} \quad F(F^{-1}(x)) = x \quad \text{for all } x.$$

- If $p = 0.5$ then $q(0.5)$ is the median of the distribution. If F is strictly increasing, then $q(0.5)$ is uniquely determined.
- In modern finance, quantiles are used to measure the risk of loss. In particular, Value at Risk (VaR) at a given level $\alpha \in (0, 1)$ represents the maximum possible loss that may occur with probability less that $1 - \alpha$ over a pre-specified time period. In other words, VaR is the $(1 - \alpha)$-quantile of the distribution of loss.

Example. For the CDF

$$F(x) = x^2, \quad x \in [0, 1],$$

calculate the median and 25*th* percentile of the corresponding distribution.

Clicker Question(s).

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Change of Variable

- Consider the transformation $Y = h(X)$ of a continuous random variable X, where $h : \mathbb{R} \to \mathbb{R}$.
  Often it is of interest to determine the distribution of the transformed variable. In particular, if X is continuous with pdf $f_X(x)$, we may want to compute the pdf of Y, $f_Y(y)$.
- A possible strategy for computing the pdf and/or CDF of $Y = h(X)$ in terms of the analogous functions for X:
  (i) Determine the range of X, and from this deduce the range of $Y = h(X)$.
  (ii) Derive the CDF of Y in terms of the cdf of X.
  (iii) If desired, differentiate the CDF of Y to obtain the pdf f_Y as a function of f_X.

Example. Suppose the pdf of X is

$$f_X(x) = \frac{1}{2}, \quad x \in (0, 2].$$

Find the pdf and CDF of $Y = \frac{1}{X}$.

Since $range(X) = (0, 2]$, $range(\frac{1}{X}) = [\frac{1}{2}, \infty)$.
To find the CDF of Y, we first need the CDF of X:

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{x}{2} & 0 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$$

Since X takes only positive values, we have for $y > 1/2$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\frac{1}{X} \leq y) = P(\frac{1}{y} \leq X) \\ &= 1 - F_X(\frac{1}{y}). \end{aligned}$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

For $y > 1/2$ we have $1/y < 2$, and hence

$$F_Y(y) = \begin{cases} 0 & y < \frac{1}{2} \\ 1 - \frac{1}{2y} & \frac{1}{2} \leq y. \end{cases}$$

From this we can find the density of Y:

$$f_Y(y) = F_Y'(y) = \frac{1}{2y^2} \quad \text{for} \;\; y \geq 1/2.$$

Exercise. For the same rv X as above, find the pdf and CDF of

$$Y = -\frac{1}{X}.$$

Example. Suppose the pdf of X is

$$f_X(x) = \frac{1}{2} \quad \text{for} \quad x \in [-1, 1].$$

Find the pdf and CDF of $Y = X^2$.

Since $range(X) = [-1, 1]$, $range(X^2) = [0, 1]$.
The cdf of X:

$$F_X(x) = \begin{cases} 0 & x < -1, \\ \frac{x}{2} + \frac{1}{2} & -1 \leq x < 1, \\ 1 & x \geq 1. \end{cases}$$

Note that the function $x \to x^2$ is not monotone over $(-1, 1)$. Therefore, for $y \in [0, 1]$, we have[2]

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

[2]See also Problem 25 in Section 8.6 (Cauchy distribution)

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

For $y \in [0,1]$, we have $\sqrt{y} \in [0,1]$ and $-\sqrt{y} \in [-1,0]$. Hence

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ \sqrt{y} & 0 \leq y \leq 1, \\ 1 & y \geq 1. \end{cases}$$

From this we can find the density of Y:

$$f_Y(y) = F_Y'(y) = \frac{1}{2} y^{-1/2}, \quad \text{for} \;\; y \in [0,1].$$

Clicker Question(s).

Expectation and Variance

Definition. If X is a continuous random variable with pdf $f(x)$ and $g : \mathbb{R} \to \mathbb{R}$ a given function, then

$$E[g(X)] := \int_{-\infty}^{\infty} g(x)f(x)dx.$$

- It follows that

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$Var(X) = E[(X - E(X))^2] = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx.$$

- Expectation and variance for continuous r.v.'s have the same general properties as for discrete r.v's.

◦ Let X and Y be either continuous or discrete r.v.'s. Then:

1. For any constants a and b
$$E(aX + bY) = aE(X) + bE(Y).$$
2. If a r.v. X is such that $a \leq X \leq b$ for two constants a and b, then
$$a \leq E(X) \leq b.$$
3. If $X \geq 0$, then
$$E(X) \geq 0.$$
4. If two random variables X and Y are defined on the same sample space S and satisfy
$$X(s) \leq Y(s) \text{ for each } \quad s \in S,$$
then
$$E(X) \leq E(Y).$$

- Let X be either continuous or discrete r.v.'s. Then:

1. $Var(X) \geq 0$.
2. For any constants a and b

$$Var(aX + b) = a^2 Var(X).$$

3. Equivalent representations of variance

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2 \end{aligned}$$

Example. Suppose X has pdf

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $E(X)$ and $Var(X)$.

Exercise. Suppose X has CDF

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{x^2}{2} & 0 \leq x < 1/2 \\ \frac{7}{4}x - \frac{3}{4} & 1/2 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Compute $E(X)$ and $Var(X)$.

Clicker Question(s).

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Selected Standard Continuous Distributions

- We will consider the following standard continuous distributions:
  - Continuous Uniform Distribution
  - Exponential Distribution
  - Normal (or Gaussian) Distribution

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Continuous Uniform Distribution

Definition. We say that a r.v. X has a uniform distribution on (a, b) if its pdf is of the form

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

This is abbreviated $X \sim U(a, b)$.

- It is easy to verify that its CDF is of the form

$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & x \in [a, b], \\ 1 & x > b. \end{cases}$$

Exercise. Show that for $X \sim U(a, b)$ we have

(i) $E(X) = \frac{a+b}{2}$

(ii) $Var(X) = \frac{(b-a)^2}{12}$

Example. Suppose that the angle measured from the x-axis to the point where the spinner stops is uniformly distributed on $[0, 2\pi]$.
You win a prize if the point lands in $[\frac{3\pi}{4}, \frac{3\pi}{2}]$. Given that the point stops in the top half of the circle, what is the probability that you win the prize.

Clicker Question(s).

Exponential Distribution

We say that X has an exponential distribution with parameter $\theta > 0$ if the density of X is

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

Shorthand notation: $X \sim Exp(\theta)$ or $X \sim Exponential(\theta)$.

The CDF of X is

$$F_X(x) = \begin{cases} 1 - e^{-x/\theta} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- This distribution can also be parameterized in terms of $\lambda = 1/\theta$, in which case

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0, \\ 0 & x \leq 0. \end{cases}$$

- Exponential distribution is used in areas such as queueing theory and reliability.

  As explained below, it also arises naturally in the context of a Poisson process.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- If X is the time to the first event of a Poisson process with the average rate of occurrence λ, then

$$X \sim Exp(1/\lambda).$$

To show this, we will find the CDF of X. For $x > 0$:

$$\begin{aligned} F(x) &= P(X \leq x) = P(\text{time to } 1^{st} \text{ occurrence} \leq x) \\ &= 1 - P(\text{time to } 1^{st} \text{ occurrence} > x) \\ &= 1 - P(\text{no occurrence in the time interval } (0, x)) \\ &= 1 - \frac{(\lambda x)^0}{0!} e^{-\lambda x} \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

Thus, $X \sim Exp(1/\lambda)$.

- This result can be generalized as follows:
  the distribution of the amount of time between the occurrences of successive events in a Poisson process is exponential[3].

Clicker Question(s).

[3] See also Problem 24 in Section 8.6.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- **Mean and variance.** Suppose that $X \sim Exp(\theta)$. Show that
$$E(X) = \theta \quad \text{and} \quad Var(X) = \theta^2.$$

We can prove these by applying integration by parts to the integrals

$$\frac{1}{\theta}\int_0^\infty xe^{-x/\theta}dx \quad \text{and} \quad \frac{1}{\theta}\int_0^\infty x^2e^{-x/\theta}dx.$$

A better way, however, is to use the gamma function.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

The **Gamma function**, $\Gamma(\alpha)$, is defined for all $\alpha > 0$ as

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

It has the following properties:

(i) $\Gamma(1) = 1$

(ii) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$.

$$\begin{aligned}\int_0^\infty x^{\alpha-1} e^{-x} dx &= -\lim_{x\to\infty} x^{\alpha-1} e^{-x} + \lim_{x\to 0} x^{\alpha-1} e^{-x} \\ &\quad +(\alpha-1)\int_0^\infty x^{\alpha-2} e^{-x} dx \\ &= (\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

(iii) If $\alpha \in \mathbb{Z}^+$, then $\Gamma(\alpha) = (\alpha - 1)!$

(iv) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

From these properties, it follows that if $X \sim exp(\theta)$, then

$$E(X) = \theta \quad \text{and} \quad E(X^2) = 2\theta^2,$$

from which we get $Var(X) = \theta^2$.

Remark. Using the Gamma function we can define the following density function

$$g(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad \text{for} \quad x \geq 0.$$

A random variable X is said to have the **Gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$ if its density function is of the above form[4].

[4] See Problem 14 in Section 8.6

Example. Suppose busses arrive according to a Poisson process with an average of 4 busses per hour.

(a) Find the probability of waiting at least 10 minutes.

(b) Find the probability of waiting at least another 10 minutes given that you have already been waiting for 5 minutes.

(a) We have $\lambda = 4/\text{hour}$, which gives

$$\theta = 1/\lambda = 1/4 \text{ hour} = 15 \text{ minutes.}$$

Let

$$X = \text{time until the bus arrives.}$$

Since $X \sim Exp(\theta = 15)$, the answer is

$$\begin{aligned} P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) &= 1 - (1 - e^{-10/15}) \\ &= e^{-10/15} = 0.513. \end{aligned}$$

General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

(b)
$$\begin{aligned} P(X > 10 + 5 | X > 5) &= \frac{P(X > 15 \text{ and } X > 5)}{P(X > 5)} \\ &= \frac{P(X > 15)}{P(X > 5)} \\ &= \frac{1 - (1 - e^{-15/15})}{1 - (1 - e^{-5/15})} \\ &= \frac{e^{-15/15}}{e^{-5/15}} = e^{5/15 - 15/15} \\ &= e^{-10/15}. \end{aligned}$$

Note
$$P(X > 15 | X > 5) = P(X > 10).$$

Exercise. Show that the exponential distribution has the "memoryless property":

$$P(X > a + b | X > a) = P(X > b) \quad \text{for any } a, b > 0.$$

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Clicker Question(s).

Computer Generation of Random Variables

- A practical problem: how to generate observations from a variable X with a given CDF F_X?
- This problem arises in the context of the Monte Carlo method, which approximates $E(X)$ using a sample mean based on simulated observations $x_1, x_2, \ldots, x_N$ from the distribution of X:

$$E(X) \approx \frac{1}{N} \sum_{i=1}^{N} x_i.$$

- A standard approach to simulation of observations from a distribution with a given CDF is the “inverse transform method”.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- **Relevant facts.** Suppose X is a continuous random variable with the CDF F_X that is strictly increasing.

  (i) Then
  $$Y := F_X(X) \sim U(0,1).$$

  (ii) Let F_X^{-1} denote the inverse function of $F_X(x)$. If $U \sim U(0,1)$, then the CDF of
  $$Y := F_X^{-1}(U)$$
  is F_X.

(i) We have

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y. \end{aligned}$$

Therefore, $Y \sim U(0, 1)$.

(ii) Similarly,

$$\begin{aligned} P(Y \leq y) &= P(F_X^{-1}(U) \leq y) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \\ &= P(U \leq F_X(y)) = F_X(y), \end{aligned}$$

where in the last line we used the fact that the CDF of U is the identity function on $(0, 1)$. Therefore

$$F_Y(y) = F_X(y).$$

- From (ii), we have the following method of generating a variate from the distribution with the CDF F:
  - generate u from $U(0,1)$
  - return $F^{-1}(u)$.

Example. For the following CDF (Arcsine distribution)

$$F(x) = \begin{cases} 0 & x \leq 0, \\ \frac{2}{\pi}\arcsin(\sqrt{x}) & 0 \leq x \leq 1, \\ 1 & x > 1, \end{cases}$$

we have $F^{-1}(u) = \sin^2(\frac{\pi u}{2})$. Thus to sample from this CDF we can use

$$X = \sin^2(\frac{\pi}{2}U),$$

or, using the identity $2\sin^2(u) = 1 - \cos(2u)$, the transformation

$$X = \frac{1}{2} - \frac{1}{2}\cos(\pi U), \quad U \sim U(0,1).$$

Exercise. Consider the CDF

$$F(x) = \begin{cases} 0 & x \leq 0, \\ 1 - e^{-x/\theta} & x > 0. \end{cases}$$

Devise a transformation h so that for $U \sim U(0,1)$ the transformed variable

$$Y := h(U)$$

has CDF F.

Clicker Question(s).

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

- Most computer software has a built-in "pseudo-random number generator" that will simulate observations from a $U(0,1)$ distribution.

  In *R* to simulate 10 values from $U(0,1)$ we use

```
> x <- runif(10)
> x
 [1] 0.5499116 0.8221231 0.9308021 0.9675079 0.1579761
 [6] 0.4402177 0.3222128 0.1695494 0.8152468 0.8040795
```

- To simulate variates from other distributions we can apply the inverse transformation.

  For example, to simulate values from *Exp*(1) we use the inverse of its CDF given by

$$F^{-1}(u) = -\ln(1-u), \quad u \in (0,1).$$

  In *R*:

```
> -log(1-x)
 [1] 0.7983113 1.7266636 2.6707855 3.4267577 0.1719468
 [6] 0.5802073 0.3889219 0.1857869 1.6887347 1.6300465
```

- The "inverse method" can also be applied to discrete distributions.

  In this case, however, we have to use the following more general definition of the inverse of a CDF:

$$F^{-1}(y) = \min\{x : \; F(x) \geq y\}.$$

Example. To generate a random value from the distribution with the following p.f.

$$f(x) = \begin{cases} p & \text{for } x = 0, \\ 1 - p & \text{for } x = 1, \end{cases}$$

we can use:

- generate u from $U(0, 1)$
- if $u < p$ return 0, otherwise return 1.

Normal Distribution

Definition. A random variable X is said to have a **Normal distribution** with mean μ and variance σ^2 if the density of X is

$$f(x) \equiv f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Shorthand: $X \sim N(\mu, \sigma^2)$.

- This distribution is also known as Gaussian distribution. The shorthand notation

$$X \sim G(\mu, \sigma)$$

means that X has Gaussian distribution with mean μ and standard deviation σ.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Figure: Carl Friedrich Gauss, 1777-1855.

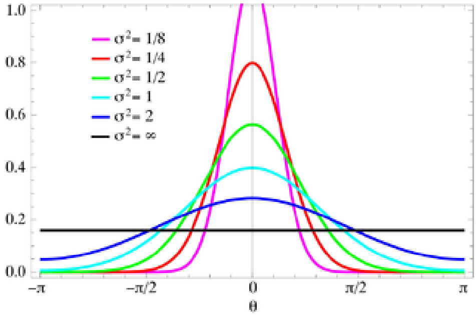


Figure: Normal densities with mean zero and increasing standard deviation.

- The shape of the normal probability density function is often termed a “bell shape” or “bell curve”.

Properties of the normal distribution.

(1) For $X \sim N(0, \sigma^2)$, the distribution is symmetric about 0:

$$f(-x) = f(x), \ x \in \mathbb{R},$$

which implies that for any x we have

$$P(X \leq -x) = P(X \geq x) = 1 - P(X \leq x). \qquad (2)$$

(2) For $X \sim N(\mu, \sigma^2)$, the distribution is symmetric about μ:

$$f(\mu - x) = f(\mu + x), \ x \in \mathbb{R},$$

which implies that for any μ and x we have

$$P(X \leq \mu - x) = P(X \geq \mu + x). \qquad (3)$$

(3) For $X \sim N(\mu, \sigma^2)$, the median is μ (follows from (3)).

(4) For $X \sim N(\mu, \sigma^2)$, the density is unimodal with the peak at μ. Thus the mode of this distribution is equal to μ.

(5) mean and variance are the parameters:

$$E(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2.$$

The first result follows from the symmetry of the distribution, while the second can be proven by using the Gamma function.

Clicker Question(s).

- Many probabilists and statisticians would argue that the normal distribution is the most important distribution.
- The main reason for this is known as the Central Limit Theorem, which we will discuss at the end of the course.

  According to this result, a sum of many independent and identically distributed random variables (continuous or discrete) has approximately normal distribution:

$$\sum_{i=1}^{N} X_i \overset{\text{distr}}{\approx} \text{Normal distribution.}$$

- Therefore the normal distribution models many different phenomena:

  - error measurements in experiments,
  - test scores on exams,
  - measurements of heights and weights in large populations, etc..

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Working with the CDF

- Major problem with the normal random distribution: if $X \sim N(\mu, \sigma^2)$, then

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = ???$$

  Functions of the form e^{-x^2} do not have elementary antiderivatives! Therefore, in practice, we have to use numerical methods.

- We can compute

$$P(a \leq X \leq b)$$

  if we know the CDF of $N(\mu, \sigma^2)$. It turns out that it suffices to know only the values of CDF for a particular set of parameters: $\mu = 0$ and $\sigma = 1$.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Definition. We say that Z is a **standard normal** random variable if $Z \sim N(0,1)$.

Frequently in probability and statistics literature, the density of the standard normal random variable is denoted

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \ x \in \mathbb{R},$$

and the CDF of a standard normal random variable is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy, \ x \in \mathbb{R}.$$

Values of the function $\Phi(x)$ are tabulated in what are called "Standard Normal Tables" or "Z-Tables".

- The table gives only values of $\Phi(x)$ for $x > 0$.

  For negative arguments use

  $$\Phi(-x) = 1 - \Phi(x), \tag{4}$$

  which follows from (2).

- To find $\Phi(x)$ for a positive x that is not listed in the table we can use interpolation.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Example. Suppose that $Z \sim N(0, 1)$. Compute

(i) $P(Z \leq 2.15)$
(ii) $P(Z \geq 2.05)$
(iii) $P(Z \leq -1.29)$
(iv) $P(-1.29 < Z \leq 2.15)$

Answers:

(i) $\Phi(2.15) = 0.98422$.

(ii)

$$P(Z \geq 2.05) = 1 - P(Z < 2.05) = 1 - \Phi(2.05) = 1 - 0.97982 = 0.02018$$

(iii) use (4):

$$P(Z \leq -1.29) = P(Z \geq 1.29) = 1 - \Phi(1.29) = 1 - 0.90147 = 0.09853$$

(iv)

$$P(-1.29 < Z \leq 2.15) = P(Z \leq 2.15) - P(Z < -1.29) = 0.98422 - 0.09853$$

which gives 0.88569.

Check:

```
> pnorm(2.15)
[1] 0.9842224

> 1-pnorm(2.05)
[1] 0.02018222

> pnorm(-1.29)
[1] 0.09852533

> pnorm(2.15) - pnorm(-1.29)
[1] 0.8856971
```

Inverse problem. For some questions we have to work with the inverse Φ^{-1} (quantiles of the distribution)

Example. Suppose that $Z \sim N(0,1)$.

(i) Find a such that $P(Z \leq a) = 0.82$
(ii) Find b such that $P(Z \leq b) = 0.25$
(iii) Find c such that $P(Z \geq c) = 0.95$
(iv) Find d such that $P(|Z| \leq d) = 0.98$

Answers:
(i) $a = \Phi^{-1}(0.82) = 0.9154$.
(ii) Note that b must be negative. Use (see (4)):

$$\Phi(b) = 1 - \Phi(-b) = 0.25,$$

which gives

$$\Phi(-b) = 0.75 \quad \Rightarrow \quad -b = \Phi^{-1}(0.75) = 0.6745.$$

Answer: $b = -0.6745$.
(iii) Note that c must be negative. Hence (by (2))

$$0.95 = P(Z \geq c) = P(Z \leq -c) = \Phi(-c) \ \Rightarrow \ -c = \Phi^{-1}(0.95) = 1.6449,$$

which gives $c = -1.6449$.

(iv) For $d > 0$ we have

$$\begin{aligned} P(|Z| \leq d) &= P(-d \leq Z \leq d) = \Phi(d) - \Phi(-d) \\ &= \Phi(d) - (1 - \Phi(d)) = 2\Phi(d) - 1. \end{aligned}$$

Thus

$$0.98 = 2\Phi(d) - 1 \ \Rightarrow \ \Phi(d) = 1.98/2,$$

which gives $d = \Phi^{-1}(0.99) = 2.3263$.

Check:

```
> qnorm(0.82)
[1] 0.9153651
> qnorm(0.25)
[1] -0.6744898
> qnorm(0.05)
[1] -1.644854
> qnorm(0.99)
[1] 2.326348
```

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Theorem. If $Y \sim N(\mu, \sigma^2)$, then for the variable

$$Z = \frac{Y - \mu}{\sigma}$$

we have

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

The operation of centering a random variable and scaling it by the reciprocal of the standard deviation is called "standardization".

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Example. For $X \sim N(2, 9)$:

(i) find
$$P(X \geq 2.5),$$

(ii) find a number c such that
$$P(X > c) = 0.05$$

Answers:

(i)
$$\begin{aligned} P(X \geq 2.5) &= P\left(\frac{X-2}{\sqrt{9}} \geq \frac{2.5-2}{\sqrt{9}}\right) \\ &= P\left(Z \geq \frac{0.5}{3}\right) = 1 - \Phi(0.1666667) \\ &= 1 - 0.5662 = 0.4338.^{5} \end{aligned}$$

[5] if we round 0.1666667 to 2 decimal places, then we can use the table: $1 - \Phi(0.17) = 1 - 0.57 = 0.43$.

(ii) For $X \sim N(2, 9)$, we need to find c such that

$$P(X > c) = 0.05.$$

We have

$$\begin{aligned} P(X > c) &= P\left(\frac{X-2}{\sqrt{9}} > \frac{c-2}{\sqrt{9}}\right) \\ &= P\left(Z > \frac{c-2}{3}\right) = 1 - \Phi\left(\frac{c-2}{3}\right). \end{aligned}$$

We want c that solves

$$1 - \Phi\left(\frac{c-2}{3}\right) = 0.05 \ \Rightarrow\ \Phi\left(\frac{c-2}{3}\right) = 0.95,$$

which gives

$$\frac{c-2}{3} = \Phi^{-1}(0.95) \ \Rightarrow\ c = 3 \times 1.6449 + 2 = 6.9346.$$

Exercise. For X as above find c such that $P(X > c) = 0.95$.

CONTINUOUS RANDOM VARIABLES (CH8)
General Facts
CDF
Quantiles
Change of Variable
Expectation and Variance
Selected Standard Continuous Distributions
Continuous Uniform Distribution
Exponential Distribution
Computer Generation of Random Variables
Normal Distribution
Working with the CDF

Exercise. Suppose the score of a randomly selected student on Midterm 2 follows a normal distribution with mean 81 and variance 36.

(i) If a student is selected at random, compute the probability that the student's score exceeds 75.

(ii) Knowing that the student's score is greater than 75, what is the probability that the score is larger than 85?

Clicker Question(s).

Definition The **68-95-99.7 Rule** states that if $X \sim N(\mu, \sigma^2)$, then[6]

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx .68$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx .95$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx .997$$

[6] see Problem 8.5.1 in the Notes.