Statistics I – Chapter 6 Continuous Probability Distributions

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 17, 2012

Introduction

- In Chapter 5, we discussed discrete probability distributions.
- In this chapter, we discuss **continuous probability distributions**.
  - Continuous distributions describe **continuous random variables**.
  - Things are **measured** rather than counted.

Road map

- **Basic concepts**
- Uniform distributions
- Exponential distributions
- Normal distributions

Probabilities for a continuous RV

- For a continuous random variable, the concept of probability should be used with cautions.
  - Let X be the temperature of this room at tomorrow noon.
  - Probably $X \in [15, 25]$.
  - What is $\Pr(X = 20)$? **Zero**!
  - Some probabilities that make sense: $\Pr(X \geq 20)$, $\Pr(18 \leq X \leq 22)$, $\Pr(X \leq 24)$, etc.
- There is a probability for a **range** of possible values.
- There is **no** probability for a **single value**!

Probability density functions

- A continuous distribution is described by a **probability density functions** (pdf).
  - Typically denoted by $f(x)$, where x is a possible value.
  - Satisfies $\int_{x \in S} f(x)dx = 1$, where S is the sample space.
  - For each possible value x, the function gives the probability **density**. It is **not** a probability!
- Recall that for a discrete distribution, we define a probability mass function.
- And the sum/integral of density becomes mass.
- So the **integral of a pdf** over a range gives probability!

Probability density functions

- Suppose a random variable X has the following pdf:

$$f(x) = kx^2 \quad \forall x \in [0, 1].$$

Let $S = [0, 1]$ be the sample space.

  - What is the value of k?
  - What is $\Pr(X \geq \frac{1}{2})$?
  - What is the **expected value** $\mathbb{E}[X]$?
  - What is the **variance** $\mathrm{Var}(X)$?

Probability density functions

- The pdf: $f(x) = kx^2$ for $x \in [0, 1]$.
- For it to really be a pdf, we need

$$\int_0^1 f(x)dx = \int_0^1 kx^2 dx = 1.$$

Why?

- So we have

$$\int_0^1 kx^2 dx = k \int_0^1 x^2 dx = k\left(\frac{1}{3}\right) = 1,$$

which implies $k = 3$.

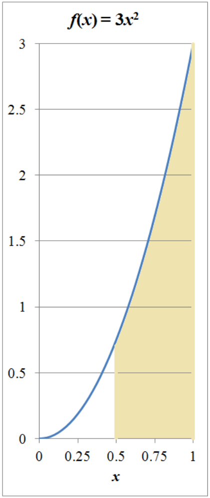
Probability density functions

- For $\Pr(X \geq \frac{1}{2})$, we have

$$\Pr\left(X \geq \frac{1}{2}\right) = \int_{\frac{1}{2}}^{1} f(x)dx = \int_{\frac{1}{2}}^{1} (3x^2)dx$$
$$= 3\left(\frac{1 - \frac{1}{8}}{3}\right) = \frac{7}{8}.$$

- For the expectation, we have

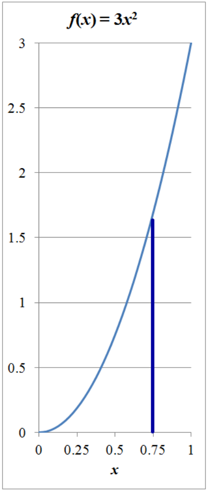
$$\mathbb{E}[X] \equiv \int_{x \in S} xf(x)dx = \int_0^1 x\left(3x^2\right)dx$$
$$= 3\int_0^1 x^3 dx = 3\left(\frac{1}{4}\right) = \frac{3}{4}.$$



Probability density functions

- The pdf: $f(x) = 3x^2$ for $x \in [0, 1]$.
- For the variance, we have

$$\begin{aligned}
\text{Var}(X) &\equiv \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \int_{x \in S} (x - \mathbb{E}[X])^2 f(x) dx \\
&= \int_0^1 \left(x - \frac{3}{4}\right)^2 (3x^2) dx \\
&= 3 \int_0^1 \left(x^4 - \frac{3}{2}x^3 + \frac{9}{16}x^2\right) dx \\
&= 3\left(\frac{1}{5} - \frac{3}{8} + \frac{3}{16}\right) = \frac{3}{80}.
\end{aligned}$$



Probability density functions

- In general, for any continuous random variable X, $\Pr(X = x) = 0$ for any single value x.
- $\Pr(X \in I)$ can be found for any interval I by doing an **integration**.
  - I may be of infinite length.

Cumulative distribution functions

- The cumulative distribution function (cdf), defined as

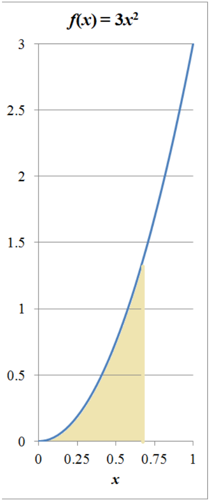
$$F(x) \equiv \Pr(X \leq x),$$

indicates the cumulative probability up to x.

- The cdf of $f(x) = 3x^2$ over $S = [0, 1]$ is

$$F(x) = \int_0^x f(y)dy = 3\int_0^x y^2 dy = 3\left(\frac{x^3}{3}\right) = x^3.$$

- In general, $\frac{d}{dx}F(x) = f(x)$.



Road map

- Basic concepts
- **Uniform distributions**
- Exponential distributions
- Normal distributions

Uniform distributions

- Sometimes the probability density of a RV is **constant**.
- In this case, we say the RV follows a **uniform distribution**:

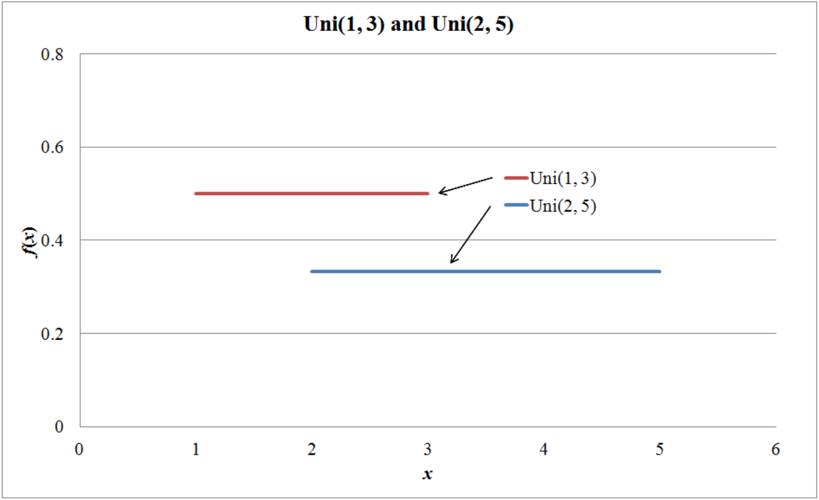
Definition 1 (Uniform distribution)

A random variable X follows the uniform distribution with lower bound $a \in \mathbb{R}$ and upper bound $b \in \mathbb{R}$, denoted by $X \sim \mathrm{Uni}(a, b)$, if its pdf is

$$f(x|a,b) = \frac{1}{b-a}$$

for all $x \in [a, b]$.

Graphing uniform distributions



Expectations and variances

- The mean and variance of $X \sim \text{Uni}(a, b)$ can be derived:

Proposition 1

Let $X \sim \text{Uni}(a, b)$, *then*

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

Proof. Homework! □

Cumulative distribution functions

- The cdf of $X \sim \text{Uni}(a, b)$ can be derived:

Proposition 2

Let $X \sim \text{Uni}(a, b)$, *then*

$$F(x|a, b) = \frac{x-a}{b-a}.$$

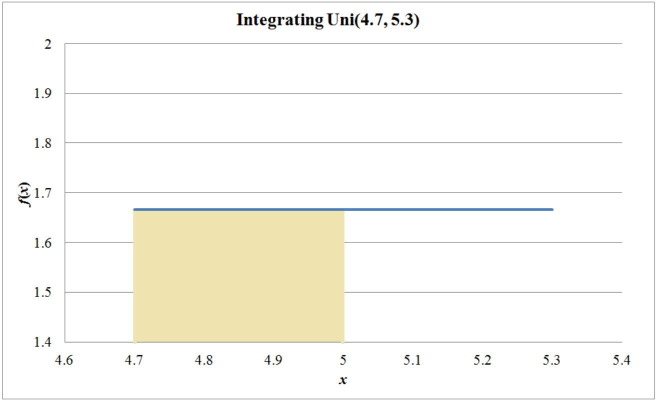
Proof. Trivial. □

An example

- Let X be the weight of a box of oranges sold at a particular price. Though ideally each box should be of 5 kg, there are some errors. Suppose X follows a uniform distribution within 4.7 kg and 5.3 kg.
  - The pdf: $f(x|4.7, 5.3) = \frac{1}{5.3-4.7} = \frac{1}{0.6} \approx 1.67$.
  - $\mathbb{E}[X] = \frac{4.7+5.3}{2} = 5$ kg.
  - $\text{Var}(X) = \frac{(5.3-4.7)^2}{12} = 0.03 \text{ kg}^2$.
  - Standard deviation ≈ 0.17 kg.

An example (cont'd)

- $X \sim \text{Uni}(4.7, 5.3)$.
  - $F(x|4.7, 5.3) = \dfrac{x - 4.7}{0.6}$.



Wait!

- How come $f(x) \approx 1.67 > 1$?
- It is a **density** function, not a probability function!
- In general:
  - For any pmf, $\Pr(x) \leq 1$ for all possible x.
  - For any pdf, $f(x)$ may be > 1 for some possible x.
  - For any cdf, $F(x) \leq 1$ for all possible x.

Remarks for uniform distributions

- For a uniformly distributed random variable, the probability density is constant.
- Except in some artificial situations, this assumption is **typically not true**, especially in natural environments.
  - A normal distribution may be a better alternative.
- Nevertheless, uniform distributions are widely used in operations research, management science, and economics due to its **tractability**.

Road map

- Basic concepts
- Uniform distributions
- **Exponential distributions**
  - Basic properties
  - Exponential and Poisson distributions
- Normal distributions

Exponential distributions

- For some situations, the probability density decreases **geometrically**.
  - Similar to radioactive decay (though it is not a probability).
- In Statistics, we are particularly interested in those functions decreasing **exponentially**.
- For example, e^{-x} for $x \in [0, \infty)$.
  - Note that

$$\int_0^\infty e^{-x} dx = -e^{-x}\Big|_0^\infty = -(0-1) = 1.$$

    So e^{-x} is indeed a pdf.

Exponential distributions

- In general, the **rate of decay** may vary.
- Let λ be the "rate", the corresponding exponential function is $e^{-\lambda x}$.
  - The **larger** the λ is, the **faster** the density decays.
- But $\int_0^\infty e^{-\lambda x} dx \neq 1$! So we need to multiply a constant λ for adjustment.

Exponential distributions

- We define the **exponential distribution** as follows:

Definition 2 (Exponential distribution)

A random variable X follows the exponential distribution with rate $\lambda \in \mathbb{R}$, denoted by $X \sim \text{Exp}(\lambda)$, if its pdf is

$$f(x|\lambda) = \lambda e^{-\lambda x}$$

for all $x \in [0, \infty)$.

- λ is the rate of decay.

Exponential distributions: Applications

- The interarrival time between two consumers at a store.
- The interarrival time between two packets at a router on a computer network.
- The service time of a consumer at a counter.
- The service time of a patient in a hospital.
- The lifetime of a product.
- The rate λ is measured as "number of occurrences per time unit".

Expectations and variances

- The mean and variance of $X \sim \text{Exp}(\lambda)$ can be derived:

Proposition 3

Let $X \sim \text{Exp}(\lambda)$*, then*

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \textit{and} \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Proof. For the expectation, we have

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \lambda \int_0^\infty x e^{-\lambda x} dx.$$

Expectations and variances

Proof (cont'd). By applying integration by parts, we have

$$\int_0^\infty xe^{-\lambda x}dx = x\left(-\frac{1}{\lambda}e^{-\lambda x}\right)\Big|_0^\infty - \int_0^\infty -\frac{1}{\lambda}e^{-\lambda x}dx.$$

For the first term, we know $\lim_{x\to 0}\frac{x}{e^{\lambda x}} = 0$, so the first term disappears. For the second term, it is

$$\frac{1}{\lambda}\int_0^\infty e^{-\lambda x}dx = -\frac{1}{\lambda^2}e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda^2}.$$

Therefore, the expectation is $\mathbb{E}[X] = \lambda(0 + \frac{1}{\lambda^2}) = \frac{1}{\lambda}$. The proof for the variance is left as homework. □

Intuitions for the expectations

- Recall that for $X \sim \text{Exp}(\lambda)$, the rate λ is measured as the number of occurrences per time unit.
  - E.g., for the arrival process of consumers into a store, if $\lambda = 5$ per hour, then in average **five** consumers enter the store **in an hour**.
- The expectation of X is $\frac{1}{\lambda}$, which is measured as "the time between two occurrences."
  - E.g., if in average five consumers enter in an hour, in average **one** consumer enters **every 12 minutes**.
  - This 12-minute interarrival time is the mean of X, which is $\frac{1}{5}$ "hour" = 12 minutes.

Cumulative distribution functions

- The cdf of $X \sim \text{Exp}(\lambda)$ can be derived:

Proposition 4

Let $X \sim \text{Exp}(\lambda)$, *then*

$$F(x|\lambda) = 1 - e^{-\lambda x}.$$

Proof. We have

$$\Pr(X < x) = \int_0^x \lambda e^{-\lambda z} dz = \lambda \left(\frac{1}{-\lambda} \right) e^{-\lambda z} \Big|_0^x = -\left(e^{-\lambda x} - 1 \right),$$

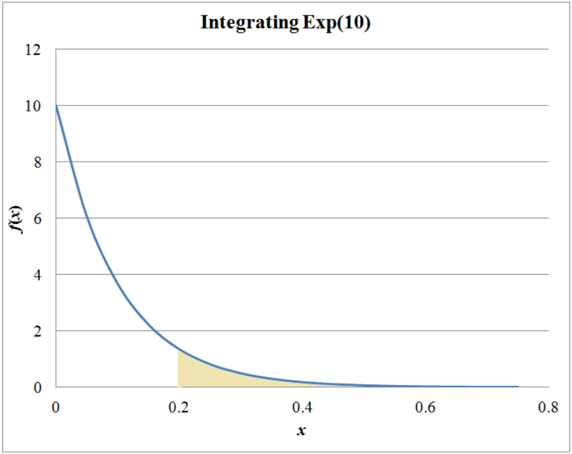
which implies $F(x|\lambda) = 1 - e^{-\lambda x}$. □

An example

- Let X be the interarrival time of buses at a particular bus stop. Suppose X follows an exponential distribution with rate 10 per hour.
  - The pdf: $f(x|10) = 10e^{-10x}$.
  - $\mathbb{E}[X] = \frac{1}{10} = 0.1$ hour $= 6$ minutes.
  - $\text{Var}(X) = \frac{1}{10}^2 = 0.01$ hour^2.
  - Standard deviation $= 0.1$ hour $= 6$ minutes.

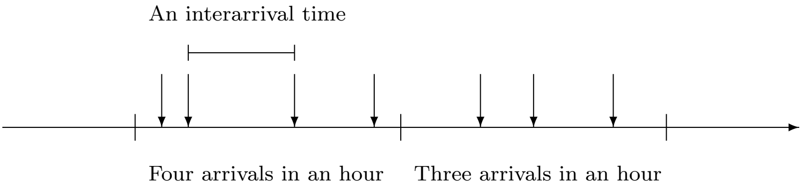
An example (cont'd)

- $X \sim \text{Exp}(10)$
  - $F(x|10) = 1 - e^{-10x}$.
  - $\Pr(X > 0.2) = 1 - F(0.2|10) = e^{-2} \approx 0.135$.



Exponential and Poisson distributions

- A Poisson RV counts the number of arrivals in a time interval.
- An Exponential RV measures the interarrival time.



- May we establish a relationship between the two distributions?

Exponential and Poisson distributions

- The following proposition connects the two distributions:

Proposition 5

Consider an arrival process within a fixed interval $[0, t]$, $t > 0$. *Let* $X \sim \text{Poi}(\lambda t)$ *be the number of arrivals, then the interarrival time* $Y \sim \text{Exp}(\lambda t)$.

Nonrigorous Proof. We may divide the interval into n pieces, i.e., $[0, \frac{t}{n})$, $[\frac{t}{n}, \frac{2t}{n})$, ..., $[(n-1)\frac{t}{n}, t]$. Let X_i be the number of arrivals in piece i, $i = 1, ..., n$. If n is large enough (i.e., approaching infinity), $X_i \sim \text{Ber}(\frac{\lambda t}{n})$ and X_i are independent. Then $\Pr(X_i = 0) = 1 - \frac{\lambda t}{n} = 1 - \Pr(X_i = 1)$.

Exponential and Poisson distributions

Nonrigorous Proof (cont'd). Now, let's consider the probability that an interarrival time is greater than t, i.e., $\Pr(Y > t)$. This means there is no arrival in $[0, t]$, or no arrival in each of the n pieces. Therefore,

$$\Pr(Y > t) = \lim_{n\to\infty} \left[\left(1 - \frac{\lambda t}{n} \right)^n \right].$$

By elementary Calculus, we have

$$\Pr(Y > t) = e^{-\lambda t}.$$

Because the cdf of an exponential RV with rate λt is $1 - e^{-\lambda t}$, we know $Y \sim \text{Exp}(\lambda t)$. □

Exponential and Poisson distributions

- Because t is arbitrary, we may have the modified version:

Proposition 6

Consider an arrival process within a fixed interval. Let $X \sim \text{Poi}(\lambda)$ be the number of arrivals, then the interarrival time $Y \sim \text{Exp}(\lambda)$.

- Intuition:
  - Poisson: **frequency** (e.g., arrivals per hour).
  - Exponential: **cycle** (e.g., hours per arrival).

Road map

- Basic concepts
- Uniform distributions
- Exponential distributions
- **Normal distributions**
  - Basic properties
  - Approximating binomial distributions

Normal distributions

- One of the most important distribution in Statistics.
- Also known as Gaussian distributions.
  - Named after Carl Friedrich Gauss.

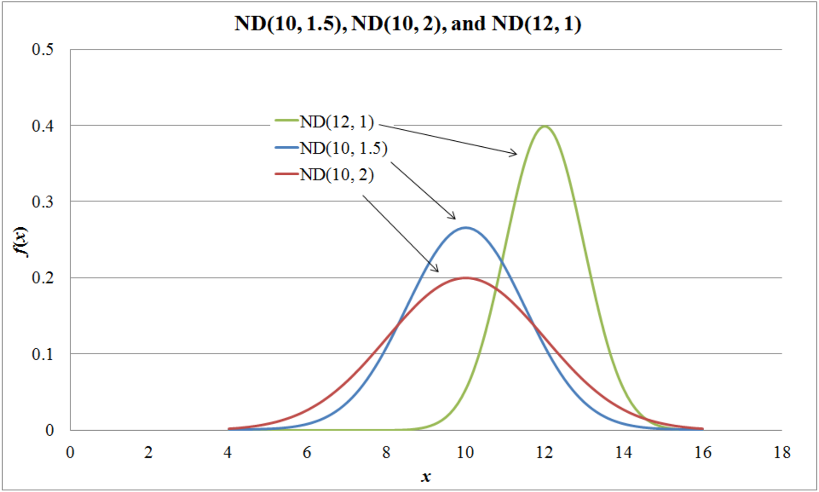
Definition 3 (Normal distribution)

A random variable X follows the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_+ = [0, \infty)$, denoted by $X \sim \mathrm{ND}(\mu, \sigma)$, if its pdf is

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

for all $x \in \mathbb{R}$.

Graphing normal distributions



Normal distributions

- A normal distribution is always **symmetric**.
  - Mean = median = mode.
  - Below (above) the mean, the probability is $\frac{1}{2}$.
- The two parameters are, by definition, its **expected value** and **standard deviation**.
  - Some researchers use the variance rather than standard deviation as the second parameter.
  - Increasing the expected value μ shifts the curve to the right.
  - Increasing the standard deviation σ makes the curve flatter.
- A normal curve is perfectly **bell-shaped**.

Normal distributions: Applications

- Natural variables: heights of people, weights of dogs, lengths of leaves, temperature of a city, etc.
- Performance: transmission time of a packet through TCP, sales made by salespeople, consumer demands, student grades, etc.
- All kinds of errors: estimation errors for consumer demand, differences from a manufacturing standard, etc.
- More importantly, some most important statistics approximately follow the normal distribution when the sample size is large enough (to be discussed in Chapter 7).

Warning!

- A normal curve spread from negative infinity to positive infinity. This is **not true** for most of the practical case!
  - E.g., student grades, heights, weights, etc.
- In using a normal distribution to approximate a practical variable, we must make sure that in our normal curve, the probability for "impossible" values to occur is **insignificant**.

Expectations, variances, and cdf

- Given a random variable $X \sim \text{ND}(\mu, \sigma)$ and its pdf $f(x|\mu, \sigma)$, we may (should) still prove that its mean and variances are indeed μ and σ^2.
  - $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\delta})^2} dx.$
  - $\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\delta})^2} dx$
- However, it is very hard if we do that from the definitions.
- The cdf of a normal curve

$$F(x|\mu, \delta) = \int_{-\infty}^{x} \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z-\mu}{\delta})^2} dz$$

does not have a closed form.

Standard normal distributions

- In general, normal distributions are useful but **hard** to use.
  - Numerically, calculating $f(x|\mu, \sigma)$ or $F(x|\mu, \sigma)$ is hard.
  - Analytically, the complicated form forbids us from deriving properties easily.
- Amazingly, all normal distributions with different parameters can have a mapping with the unique **standard normal distribution**.
  - The standard normal distribution, typically denoted as $\phi(x)$, is a normal distribution with $\mu = 0$ and $\sigma = 1$.
- Let's see how to construct the mapping.

Standard normal distributions

- Consider a random variable $X \sim \text{ND}(\mu, \sigma)$.
- Define $Z = \frac{X-\mu}{\sigma}$. Z is another random variable.
- Then $Z \sim \text{ND}(0, 1)$!

Proposition 7

If $X \sim \text{ND}(\mu, \sigma)$, *then* $Z = \frac{X-\mu}{\sigma} \sim \text{ND}(0, 1)$.

Proof. Later in the semester. □

Standard normal distributions

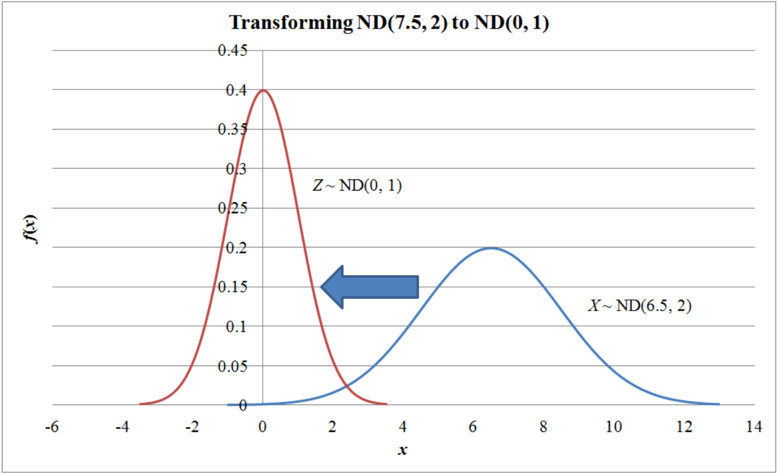
- For a value x, recall that its **z-score** is $\frac{x-\mu}{\sigma}$.
- Therefore, the standard normal distribution is sometimes called the **z distribution**.
- People has constructed the cumulative probability table for the standard normal distribution.
  - Table A.5, 6.2, or the one inside the cover in the textbook.
- Problems regarding a normal distribution with $\mu \neq 0$ and $\sigma \neq 1$ can be solved by transforming to the standard normal distribution.

Using the standard normal distribution

- Let X be a randomly selected student's score in an exam. Suppose for this exam, the mean is 6.5 (out of 10), the standard deviation is 2, and the scores are approximately normally distributed.
  - What is the probability that X is above 10 or below 0?
  - What is the probability that X is larger than 8?
  - What is the percentile that maps to a 8-point score?

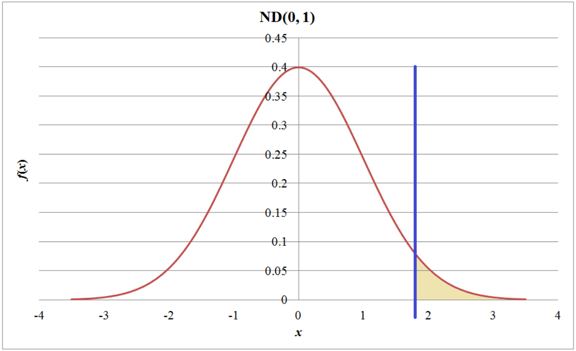
Using the standard normal distribution

- Let $Z = \frac{X-6.5}{2}$.



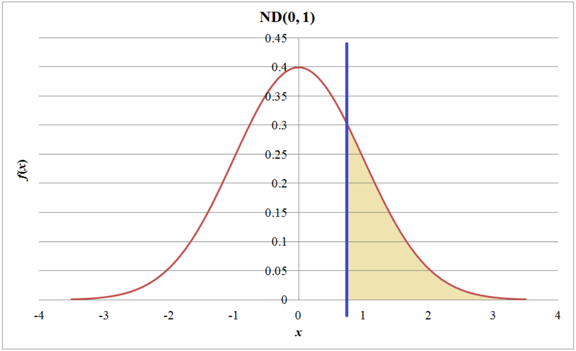
Using the standard normal distribution

- What is the probability that X is above 10 or below 0?
  - $\Pr(X \geq 10) = \Pr\left(\frac{X-6.5}{2} \geq \frac{10-6.5}{2}\right) = \Pr(Z \geq 1.75) \approx 0.04.$
  - $\Pr(X \geq 0) = \Pr\left(\frac{X-6.5}{2} \leq \frac{0-6.5}{2}\right) = \Pr(Z \leq -3.25) \approx 0.$



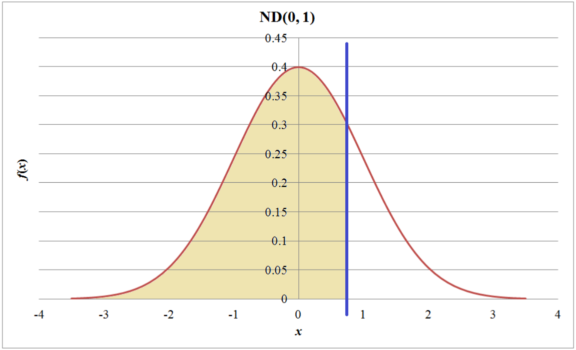
Using the standard normal distribution

- What is the probability that X is larger than 8?
  - $\Pr(X \geq 8) = \Pr(\frac{X-6.5}{2} \geq \frac{8-6.5}{2}) = \Pr(Z \geq 0.75) \approx 0.227.$



Using the standard normal distribution

- What is the percentile that maps to a 8-point score?
  - $\Pr(X \geq 8) \approx 0.227$. So $\Pr(X \leq 8) \approx 0.773$.
  - So it's around 77^{th} or 78^{th} percentile.



Standard normal distributions

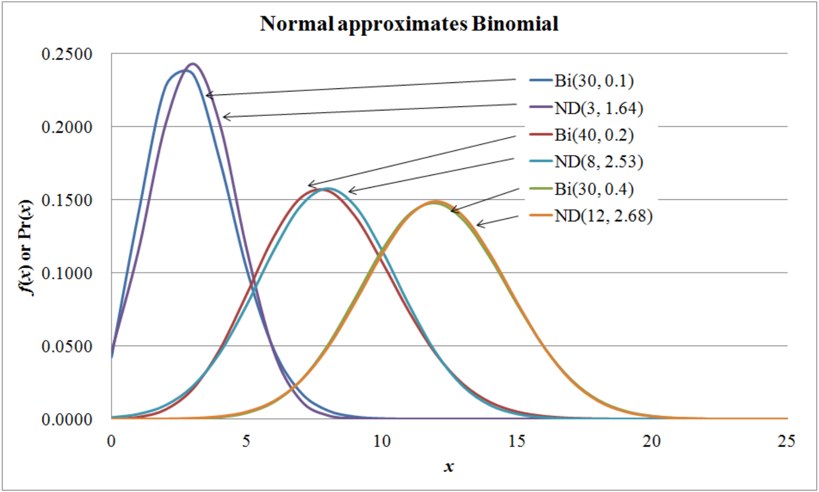
- So with
  - the **transformation** to the standard normal distribution and
  - the **probability** table of standard normal distribution,

  we are able to solve normal distribution problems regarding any values of μ and σ.
- But with MS Excel or other software, we may solve those problems directly.
- Nevertheless, we will see that the transformation plays an important role in deriving some analytical properties of inferential Statistics.

Approximating binomial with normal

- Let $X \sim \text{Bi}(n, p)$.
- When $n \to \infty$, $p \to 0$, and $np = \lambda$, $\text{Bi}(n, p) \to \text{Poi}(\lambda)$.
- So a Poisson RV can approximate a binomial RV.
- A normal RV can also approximate a binomial RV.
- When n **is large** and p **is moderate** (not close to 0 or 1), $\text{Bi}(n, p) \approx \text{ND}\left(np, \sqrt{np(1-p)}\right)$.
  - A rule of thumb: $n \geq 25$, $np > 5$, and $n(1-p) > 5$.

Approximating binomial with normal



Approximating binomial with normal

- Why $\text{Bi}(n, p) \approx \text{ND}\left(np, \sqrt{np(1-p)}\right)$?
  - A binomial distribution is always bell-shaped.
  - A binomial distribution's mean is always around its mode.
  - When $np > 5$, the mean is "far" from zero and the distribution looks like symmetric.

Approximating binomial with normal

- *Question*: Suppose we toss a fair coin 50 times. What is the probability of getting 25 to 30 heads?
- Let X be the number of heads out of the 50 trials, then $X \sim \text{Bi}(50, 0.5)$. Then we have

$$\Pr(25 \leq X \leq 30) = \sum_{x=25}^{30} \Pr(X = x) = 0.4967.$$

  as an **exact** answer.
- Let Y be the normal RV that approximates X. We know $Y \sim \text{ND}(25, 3.54)$, so

$$\Pr(25 \leq Y \leq 30) \approx \Pr(0 \leq Z \leq 1.414) \approx 0.4214.$$

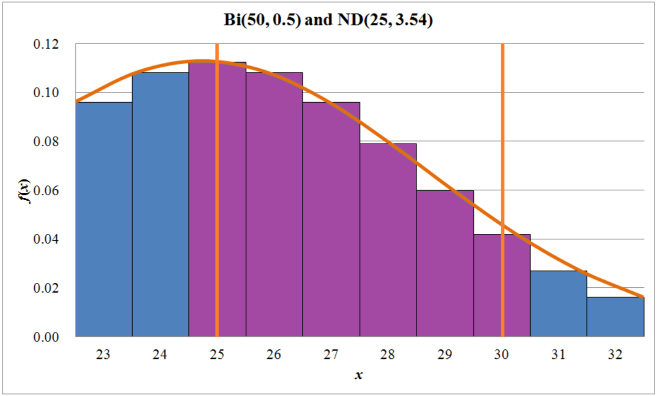
  is an **approximation**. Acurate?

Correction of continuity

- The previous approximation is not accurate because we have one more thing to do, the **correction of continuity**.
- Consider the previous example: Tossing a fair coin 50 times.
- What is the probability of getting **exactly** 20 heads?
  - Calculating based on the binomial distribution, we know the probability is positive.
  - But approximating based on the normal distribution, we will get $\Pr(Y = 20)$, which is **zero**!
  - Do not forget that the normal distribution is **continuous**.

Why correction of continuity?

- Use $Y \sim \text{ND}(25, 3.54)$ to approximate $X \sim \text{Bi}(50, 0.5)$.
  - $\Pr(25 \leq X \leq 30)$: Purple area.
  - $\Pr(25 \leq Y \leq 30)$: Area below the orange curve over $[25, 30]$.



Why correction of continuity?

- $X \sim \text{Bi}(50, 0.5)$ and $Y \sim \text{ND}(25, 3.54)$.
  - $\Pr(25 \leq Y \leq 30)$ **underestimates** $\Pr(25 \leq X \leq 30)$.
  - How to fix it?
  - $\Pr(24.5 \leq Y \leq 30.5)$!
  - $\Pr(24.5 \leq Y \leq 30.5) \approx \Pr(-0.141 \leq Z \leq 1.556) \approx 0.4963$, which is close to $\Pr(25 \leq X \leq 30) \approx 0.4967$.

Correction of continuity

- *Question*: What is the probability of getting 27 heads?
- *An exact answer*: $X \sim \text{Bi}(50, 0.5)$:

$$\Pr(X = 27) \approx 0.09596.$$

- *Approximation*: $Y \sim \text{ND}(25, 3.54)$:

$$\Pr(Y = 27) = 0,$$

but

$$\Pr(26.5 \leq Y \leq 27.5) \approx 0.09593.$$

Summary

- It is suitable for most of natural (and artificial) environments.
- The normal distribution with mean zero and standard deviation one is called the standard normal distribution.
- It approximates the binomial distribution.
- It is also the (approximate) distribution of some important statistics (to be introduced in Chapter 7).

Relations among distributions