BAYESIAN STATISTICS

Chapter 3

Instructor: Seokho Lee

Hankuk University of Foreign Studies

3. Probabilities and Distributions

3.1. Probability

3.1.1. Sample space and event

Probability numerically measures the possibilities that a certain event happens under the uncertainty. To define the probability rigorously, some concepts are needed, for example, sample space and event.

Sample space is the set of all possible outcomes in the statistical experiment or the random experiment.

Event is a subset of the sample space, consisting of any possible outcomes. Especially, the event of a single outcome is called the **simple event**.

- $A \cup B$: union
- $A \cap B$: intersection
- A^c : the complement A
- If $A \cap B = \emptyset$, A and B are disjoint

3.1.2. Definitions and properties of probabilities

Definition (Classical definition of the probability)

If all events are equally probable, then the probability of the event A is defined by

$$\Pr(A) = \frac{\text{the number of outcomes in } A}{\text{the number of all outcomes in the sample space}}$$

- This definition can be applied to the sample space that is countable and finite.
- We cannot define the probability of events with the above definition in many cases where (1) the sample space is not finite or (2) the outcomes are not assumed to be equally probable:
  - the probability of a defective in a plant
  - the probability that KOSPI will go down below 1000 points tomorrow

Definition (Axioms of probability (Kolmogorov))

We call $\Pr(\cdot)$ the probability if $\Pr(\cdot)$ satisfies

1. For any event A in the sample space $\mathcal{S}$, $0 \leq \Pr(A) \leq 1$
2. $\Pr(\mathcal{S}) = 1$
3. For the disjoint events, $A_1, A_2, \ldots$, $\Pr(A_1 \cup A_2 \cup \cdots) = \Pr(A_1) + \Pr(A_2) + \cdots$.

Some properties:

1. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
2. $\Pr(A^c) = 1 - \Pr(A)$
3. (Additivity) For n disjoint events $A_1, A_2, \ldots, A_n$,

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n)$$

4. If $A \subset B$ then $\Pr(A) \leq \Pr(B)$

3.2. Conditional probability

In a die tossing, when we know the outcome is an even number, the probability of 2 is not 1/6 anymore. The information that the outcome is even reduces the sample space $\{1, 2, 3, 4, 5, 6\}$ into $\{2, 4, 6\}$. Thus, from the classical definition of the probability, the probability of 2 becomes 1/3.

The probability of B (for example, $\{2\}$) given the fact that an event A (for example, $\{2, 4, 6\}$) occurs is called the **conditional probability** and denoted by $\Pr(B|A)$.

Definition (Conditional probability)

The conditional probability of B given an event A is

$$\Pr(B|A) = \begin{cases} \frac{\Pr(B\cap A)}{\Pr(A)} & \text{if } \Pr(A) > 0 \\ 0 & \text{if } \Pr(A) = 0 \end{cases}$$

From the above definition,

$$\Pr(A \cap B) = \Pr(B|A)\,\Pr(A).$$

This property is called **multiplication rule**.

Example (3-1)

Suppose that 45% students in a middle school commute school by walk, and 15% students are the first-graders and commute school by walk. When we meet a student going to school by walk, what is the probability that (s)he is a first-grader.

(solution) Let A be the event that a student commutes school by walk and B be the event that a student is a first-grader. Then, $\Pr(A) = 0.45$ and $\Pr(A \cap B) = 0.15$. So

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} = \frac{0.15}{0.45} = 1/3$$

Note that $\Pr(B|A) = \Pr(B)$ implies that the event A does not affect the probability of B. This is equivalent to $\Pr(A \cap B) = \Pr(A)\Pr(B)$. If such case, A and B are **mutually independent**.

3.3. Bayes theorem

For any two events A and B, we can easily show that

$$\begin{aligned}\Pr(B) &= \Pr(A \cap B) + \Pr(A^c \cap B) \\ &= \Pr(A)\Pr(B|A) + \Pr(A^c)\Pr(B|A^c).\end{aligned}$$

Definition (Law of total probability)

Suppose that the disjoint events $A_1, A_2, \cdots, A_n$ of the sample space $\mathcal{S}$ satisfying

1. $\mathcal{S} = A_1 \cup A_2 \cup \cdots \cup A_n$
2. $\Pr(A_i) > 0$, $i = 1, 2, \cdots, n$.

Then, for any event B

$$\Pr(B) = \Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2) + \cdots + \Pr(A_n)\Pr(B|A_n)$$

Bayes theorem plays an important role in Bayesian statistics.

Theorem (Bayes theorem)

Suppose that the disjoint events $A_1, A_2, \cdots, A_n$ *of the sample space* $\mathcal{S}$ *satisfying*

1. $\mathcal{S} = A_1 \cup A_2 \cup \cdots \cup A_n$
2. $\Pr(A_i) > 0$, $i = 1, 2, \cdots, n$.

Then, for any event B *of* $\Pr(B) > 0$,

$$\begin{aligned} \Pr(A_i|B) &= \frac{\Pr(A_i)\Pr(B|A_i)}{\Pr(B)} \\ &= \frac{\Pr(A_i)\Pr(B|A_i)}{\Pr(A_1)\Pr(B|A_1) + \cdots + \Pr(A_n)\Pr(B|A_n)} \end{aligned}$$

for $i = 1, 2, \cdots, n$.

Example (3-2)

The company hires 5% of the candidates. Among the candidates who are hired, 92% are graduates from colleges or higher and 8% are graduates from high schools. Among the candidates who are not hired, 3% are graduates from colleges or higher and 97% are graduates from high schools. What is the probability that the college graduate "K" will be hired by this company?

(solution) Let

$$\begin{aligned} A &= \text{a candidate is hired} \\ B &= \text{a candidate is a college graduate.} \end{aligned}$$

We know that $\Pr(A) = 0.05$, $\Pr(B|A) = 0.92$, $\Pr(B^c|A) = 0.08$, $\Pr(B|A^c) = 0.03$, and $\Pr(B^c|A^c) = 0.97$. Using the Bayes theorem, the probability $\Pr(A|B)$ is

$$\begin{aligned} \Pr(A|B) &= \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A)\Pr(B|A) + \Pr(A^c)\Pr(B|A^c)} \\ &= \frac{(0.05)(90.92)}{(0.05)(0.92) + (0.95)(0.03)} \approx 0.62. \end{aligned}$$

Thus, with the additional information for a candidate's education, a candidate "K" has 62% chance of getting a job, which is higher than the prior 5% chance of hiring.

Example (3-3)

A factory has 3 machines. Machine 1 produces 30% of the total output. And Machine 2 and Machine 3 produce 25% and 45% respectively. The defective rates are 5% for Machine 1, 3% for Machine 2, and 2% for Machine 3. Suppose we randomly draw a sample from the products and this is defective. Then what is the probability that this random sample is from Machine 3?

(solution) Let, for $i = 1, 2, 3$,

$$A_i = \text{The product is from Machine } i.$$

Then, $\Pr(A_1) = 0.30$, $\Pr(A_2) = 0.25$, and $\Pr(A_3) = 0.45$. And let

$$B = \text{The product is defective.}$$

Then,

$$\Pr(B|A_1) = 0.05, \quad \Pr(B|A_2) = 0.03, \quad \Pr(B|A_3) = 0.02.$$

From the Bayes theorem,

$$\begin{aligned} \Pr(A_3|B) &= \frac{\Pr(A_3)\Pr(B|A_3)}{\sum_{i=1}^{3}\Pr(A_i)\Pr(B|A_i)} \\ &= \frac{(0.45)(0.02)}{(0.30)(0.05) + (0.25)(0.03) + (0.45)(0.02)} \approx 0.29. \end{aligned}$$

3.4. Random variables

The output from the statistical experiments or surveys are *stochastic* rather than *deterministic*. Thus, it is natural to use the probabilistic concepts to understand its structural properties. If we are able to define a real function from a set of events to a real line, the stochastic properties of experiments are easily understood.

Consider a coin-tossing experiment. If we assign 1 to Head and 0 to Tail, then the sample space becomes $\mathcal{S} = \{0, 1\}$. We can define

$$X = \begin{cases} 1 & \text{if Head} \\ 0 & \text{if Tail} \end{cases}$$

and

$$\Pr(\text{Head}) = \Pr(X = 1) = 1/2, \quad \Pr(\text{Tail}) = \Pr(X = 0) = 1/2.$$

This variable X is not determined, but stochastically variable. A **random variable** is a function that assigns a real value to an experimental outcome appearing with probability.

3.5. Probability distribution

The **probability distribution** or **distribution** is the probabilities corresponding to the value of a random variable. For example, X that was defined in the previous page has a distribution that

x	0	1	total
p(x)	1/2	1/2	1

where $p(x) = \Pr(X = x)$.

As the above case, $p(x)$ is the probability that X takes x. We call it **probability mass function** or **pmf**. Thus, the probability mass function can describe the distribution of a discrete random variable. Any probability mass function should satisfy the following conditions:

1. $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$
2. $\Pr(a \leq X \leq b) = \sum_{a \leq x \leq b} p(x)$.

For continuous random variables (e.g., the lifetime of the electronics), it is impossible to describe the probability distribution with the probability mass function. Instead of using the point mass, the probability is defined on the interval.

For example, suppose X represents the life time of a battery. Then X is a continuous random variable. The probability that a battery runs out between 45 and 54 hours is $\Pr(45 \leq X \leq 54)$.

If we can assign to $\Pr(a \leq X \leq b)$ the area under a relevant function $f(x)$ over the interval $[a, b]$, then $f(x)$ is called the **probability density function** or **pdf**. A probability density function should satisfy the following conditions:

1. $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$,
2. $\Pr(a \leq X \leq b) = \int_a^b f(x)dx$.

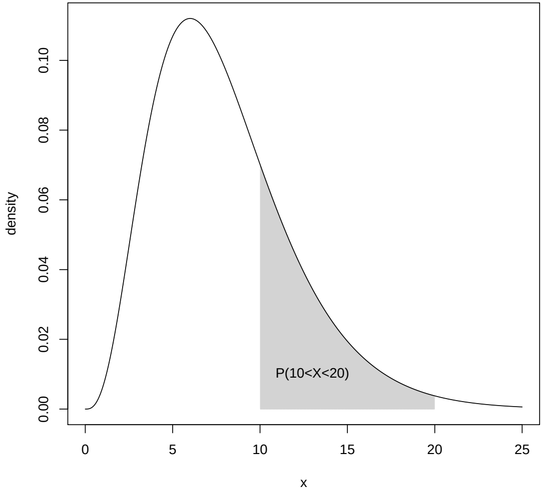


Figure: Probability density function of a continuous random variable

3.6. Joint probability distribution and marginal probability distribution

Sometimes we have two or more random variables under the probabilistic relation. In such case, it is difficult to grasp the relations and properties between them through the distribution of each variable. Thus, we need the concept of the joint probability distribution.

A **joint probability distribution** is a function from two sets A and B on a real line to the probability $\Pr(X \in A, Y \in B)$.

For the discrete random variables, X and Y, the **joint probability mass function** p_{XY} provides

$$p_{XY}(x, y) = \Pr(X = x, Y = y).$$

And for the continuous random variables, X and Y, the **joint probability density function** f_{XY} is nonnegative and satisfies

$$\Pr(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx.$$

A joint probability mass function for X and Y should satisfies

1. $p_{XY}(x,y) \geq 0$ and $\sum_x \sum_y p_{XY}(x,y) = 1$,
2. $\Pr(a \leq X \leq b, c \leq Y \leq d) = \sum_{a \leq x \leq b} \sum_{c \leq y \leq d} p_{XY}(x,y)$.

A joint probability density function for X and Y should satisfies

1. $f_{XY}(x,y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$,
2. $\Pr(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)dydx$.

The **marginal probability distribution** is obtained by summing (integrating) out the joint probability distribution along the other variables.

For the discrete case, the **marginal probability mass function** is

$$p_X(x) = \sum_y p_{XY}(x,y), \quad p_Y(y) = \sum_x p_{XY}(x,y)$$

and for the continuous case, the **marginal probability density function** is

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

Example (3-4)

The persons A and B participate in the die games. Both of them should pay the entry fee \$100 to start the game. Game 1 in that A participates gives nothing for 1 and 2, \$100 for 3 and 4, \$200 for 5 and 6. Game 2 in that B participates gives nothing for even numbers and \$100 for 1, \$200 for 3, \$300 for 5. Give the probability distribution for the sum of two persons' gain.

(solution) Let X be the gain for A and Y be the gain for B. The probability distributions for X and Y are following:

x	-100	0	100	total
$p_X(x) = \Pr(X = x)$	1/3	1/3	1/3	1

y	-100	0	100	200	total
$p_Y(y) = \Pr(Y = y)$	1/2	1/6	1/6	1/6	1

The joint distribution of X and Y is

$x \backslash y$	-100	0	100	200	total
-100	1/6	1/6	0	0	1/3
0	1/6	0	1/6	0	1/3
100	1/6	0	0	1/6	1/3
total	1/2	1/6	1/6	1/6	1

Now let $Z = X + Y$. Then Z can take the values -200, -100, 0, 100, 300. And its probability distribution is

z	-200	-100	0	100	300	total
$p_Z(z) = \Pr(Z = z)$	1/6	2/6	1/6	1/6	1/6	1

Suppose the joint distribution of X and Y is $f_{XY}(x, y)$, and the marginal distributions of X and Y are $f_X(x)$ and $f_Y(y)$ respectively. If they satisfy

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

for all x and y, then two random variables X and Y are **mutually independent**.

Example (3-4 continue)

Are X and Y independent?

(solution) X and Y are not independent because

$$p_{XY}(-100, 0) \neq p_X(-100)p_Y(0).$$

3.7. Conditional probability distribution

The **conditional probability distribution** is given as

$$f_{X|Y} = \frac{f_{XY}(x,y)}{f_Y(y)} = \begin{cases} \dfrac{\Pr(X = x, Y = y)}{\sum_u \Pr(X = u, Y = y)} & \text{if } X \text{ and } Y \text{ are discrete} \\ \dfrac{f_{XY}(x,y)}{\int_{-\infty}^{\infty} f_{XY}(u,y)du} & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

Example (3-4 continue)

What is the probability that A's gain is $0 when B gains $100?

(solution)

$$\Pr(X = 0|Y = 100) = p_{X|Y}(0|100) = \frac{p_{XY}(0,100)}{p_Y(100)} = \frac{1/6}{1/6} = 1.$$

3.8. Expectation and variance

Expectation is the expected value of a random variable on average. Expectation is denoted by $E(X)$ and defined as

$$E(X) = \begin{cases} \sum_x xp(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

Variance is a measure of the spread of a random variable from its expectation. Variance is denoted by $var(X)$ and defined as

$$var(X) = E[(X - E(X))^2] = \begin{cases} \sum_x \{x - E(X)\}^2 p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} \{x - E(X)\}^2 f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

One may use a more handy formula for variance:

$$var(X) = E(X^2) - \{E(X)\}^2.$$

Example (3-4 continue)

What are the variances of X, Y, and Z?

(solution) We can easily verify that $E(X) = E(Y) = E(Z) = 0$. Thus,

$$var(X) = \{(-100)^2(\tfrac{1}{3}) + (0^2)(\tfrac{1}{3}) + (100^2)(\tfrac{1}{3})\} - 0^2 = \tfrac{20000}{3}.$$

In a similar manner, we get $var(Y) = 40000/3$ and $var(Z) = 80000/3$.

For the random variables X, Y and the constant a, b, c, the expectation and variance has the following properties:

1. $E(aX + bY + c) = aE(X) + bE(Y) + c$,
2. $var(aX + b) = a^2 var(X)$, and
3. $var(aX + bY + c) = a^2 var(X) + b^2 var(Y) + 2ab cov(X, Y)$. If X and Y are independent, then $var(aX + bY + c) = a^2 var(X) + b^2 var(Y)$.

Example (3-4 continue)

Let U be the total gain when both of A and B participate in Game 1, V be the total gain when both participate in Game 2, and W be the total gain when A participates in Game 1 and B in Game 2. Compute the expectations and variances of U, V, and W.

(solution) Note that $U = 2X$, $V = 2Y$, and $W = Z = X + Y$. Thus,

$$\begin{aligned} E(U) &= 2E(X) = 0 \\ E(V) &= 2E(Y) = 0 \\ E(W) &= E(Z) = 0 \\ var(U) &= 2^2 var(X) = \frac{80000}{3} \\ var(V) &= 2^2 var(Y) = \frac{160000}{3} \\ var(W) &= var(Z) = \frac{80000}{3} \end{aligned}$$

3.9. Discrete probability distributions

3.9.1. Binomial distribution

Bernoulli trial is the random experiment where the result is one of the exclusive two outcomes, success (S) and failure (F). For example, a coin tossing (Head/Tail), smoking status (Yes/No), and problem solving (True/False), etc. are in such case.

Bernoulli distribution is the probability distribution for a random variable X taking a value 1 for a success with the success probability p and 0 for a failure with probability $1 - p$. We denote $X \sim \text{Bernoulli}(p)$ and its probability mass function is

$$\Pr(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1.$$

Binomial distribution is the probability distribution for a random variable X which is the number of successes from n independent Bernoulli trials with a success probability p. We denote $X \sim \mathrm{B}(n, p)$ and its probability mass function is

$$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

$\text{Bernoulli}(p)$ is a special case of $\mathrm{B}(n, p)$ with $n = 1$.

For $X \sim \mathrm{B}(n, p)$, $E(X) = np$ and $var(X) = np(1 - p)$.

Example (3-5)

Suppose we toss five times a unfair coin having a probability 1/3 for a head. Let X be the number of heads over five trials. (1) What is the probability distribution for X? (2) Compute $E(X)$ and $var(X)$. (3) Compute the probability that the head appears at least 3 times.

(solution) Since $X \sim \mathrm{B}(5, 1/3)$, its probability mass function is

$$\Pr(X = x) = \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x}, \quad x = 0, 1, 2, 3, 4, 5.$$

Thus, its expectation and variance are

$$E(X) = np = (5)(1/3) = 5/3 \approx 1.667,$$

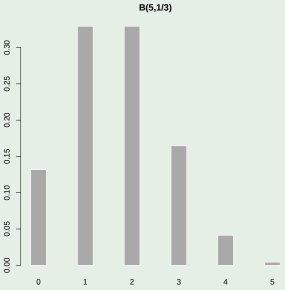
$$var(X) = np(1-p) = (5)(1/3)(2/3) = 10/9 \approx 1.111.$$

And

$$\begin{aligned} \Pr(X \geq 3) &= 1 - \Pr(X \leq 2) = 1 - \sum_{x=0}^{2}\binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x} \\ &\approx 1 - 0.13169 - 0.32922 - 0.32922 = 0.20987. \end{aligned}$$

Example (3-5 continue)

```
> n <- 5; p <- 1/3; x <- 0:n
> px <- dbinom(x,n,p)
> barplot(px, space=2, names.arg=x,
col='darkgray', border='white',
main='B(5,1/3)')
> dev.copy2pdf(file='fig3-2.pdf')
quartz
     2
> 1-pbinom(2,n,p)
[1] 0.2098765
```



3.9.2. Poisson distribution

Poisson distribution is for the number of events occurring on a time interval or an area. The Poisson distribution is modeled for rare events; for example, 'the number of accidents in a certain area during a day', 'the number of typographical errors per a page', and 'the number of military engagements in a year.'

Let X be the number of events occurs per a unit time (or area) when the events happens on average λ times per a unit time (or area). Then we denote $X \sim \text{Poisson}(\lambda)$ and its probability mass function is

$$\Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots .$$

We can verify that $E(X) = \lambda$ and $var(X) = \lambda$.

Example (3-6)

In a certain region, there are 0.5 traffic accidents during a day on average. Suppose X is the random variable representing the number of accidents during a day and assume that X follows Poisson distribution.

1. *What is the probability that there is no accident during a day? And what is the probability that there are at least 3 accidents during a day?*
2. *Compute the expectation and variance for the number of accidents during a day.*

(solution)

1. Since $X \sim \text{Poisson}(0.5)$, the probability mass function of X is

$$\Pr(X = x) = \frac{e^{-.5}(.5)^x}{x!}, \quad x = 0, 1, 2, \ldots.$$

   Thus, the probability of no accident is

$$\Pr(X = 0) = \frac{e^{-.5}(.5)^0}{0!} = 0.607$$

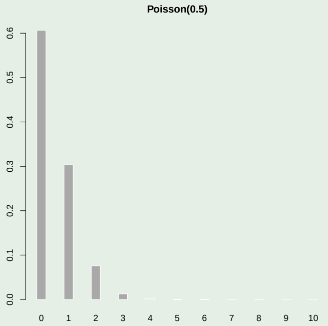
   and the probability of ≥ 3 accidents is

$$\begin{aligned} \Pr(X \geq 3) &= 1 - \Pr(X \leq 2) = 1 - \sum_{x=0}^{2} \frac{e^{-.5}(.5)^x}{x!} \\ &= 1 - 0.986 = 0.014. \end{aligned}$$

2. $E(X) = var(X) = \lambda = 0.5.$

Example (3-6 continue)

```
> dpois(0,0.5)
[1] 0.6065307
> 1-ppois(2,0.5)
[1] 0.01438768
>
> lambda<-0.5
> x<-0:10
> px<-dpois(x,lambda)
> barplot(px,space=2,names.arg=x,
col='darkgray',border='white',
main='Poisson(0.5)')
> dev.copy2pdf(file='fig3-3.pdf')
```



It is known that Poisson(np) approximates B(n, p) well when n is large, p is small, and $np \leq 10$.

Example (3-7)

In a certain factory, the defective rate in a production line is known to be 0.0005. Let X be the number of defectives among 100 random samples from this production line. Assume that X follows the binomial distribution.

1. *What is the probability that there are at least 3 defectives in this sample?*
2. *Compute the same probability using Poisson approximation.*

(solution)

1. Since $X \sim \mathrm{B}(100, 0.0005)$, the probability of ≥ 3 defectives is

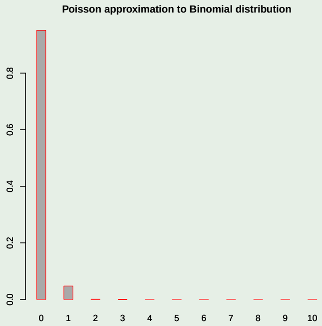
$$\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - \sum_{x=0}^{2} \binom{100}{x}(.0005)^x(.9995)^{100-x} \approx .00002.$$

2. Approximately, $X \sim \mathrm{Poisson}(0.05)$. Thus,

$$\Pr(X \geq 3) = 1 - \Pr(X \leq 2) \approx 1 - \sum_{x=0}^{2} \frac{e^{-.05}(.05)^x}{x!} \approx .00002.$$

Example (3-7 continue)

```
> 1-pbinom(2,100,0.0005)
[1] 1.949120e-05
> 1-ppois(2,0.05)
[1] 2.006749e-05
>
> n<-100; p<-0.0005; x<-0:10
> bx<-dbinom(x,n,p)
> barplot(bx,space=2,names.arg=x,
col='darkgray',border='white',
main='Poisson approximation to
Binomial distribution')
> lambda<-n*p
> px<-dpois(x,lambda)
> par(new=TRUE)
> barplot(px,space=2,names.arg=x,
col="#FFFFFF00",border='red')
> dev.copy2pdf(file='fig3-4.pdf')
```



3.10. Continuous probability distributions

3.10.1. Normal distribution

The probability density function for a normal random variable X with mean μ and standard deviation σ is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

In short, we denote $X \sim \mathrm{N}(\mu, \sigma^2)$. The density functions is bell-shaped, centered around μ. μ is a center of the distribution and σ is a measure of dispersion.

When $\mu = 0$ and $\sigma = 1$, $N(0,1)$ is called the **standard normal distribution**. If $X \sim \mathrm{N}(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$. This transformation is called **standardization**.

Note that

$$\Pr(a < X < b) = \Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi(\tfrac{a-\mu}{\sigma}) - \Phi(\tfrac{b-\mu}{\sigma}),$$

where $\Phi(c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, the cumulative distribution function for the standard normal distribution.

Example (3-8)

500 high school students participate in a math competition. The test score follows the normal distribution of the mean 125 and the standard deviation 25.

1. *How many students will get the score between 150 and 180?*
2. *Which score is the threshold for top 5%?*

(solution)

1. Since $X \sim \mathrm{N}(125, 25^2)$, we know $Z = \frac{X-125}{25} \sim \mathrm{N}(0,1)$. The probability of $150 < X < 180$ is

$$\begin{aligned} \Pr(150 < X < 180) &= \Pr(\tfrac{150-125}{25} < Z < \tfrac{180-125}{25}) = \Pr(1 < Z < 2.2) \\ &= \Phi(2.2) - \Phi(1) \approx .1448. \end{aligned}$$

   Therefore, $.1448 \times 500 \approx 72$ students will get the target score.

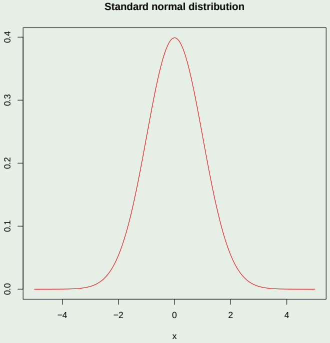
2. Note that

$$\Pr(X \geq x) = \Pr(Z \geq \tfrac{x-125}{25}) = .05.$$

   From the standard normal table, $\Pr(Z \geq 1.64) \approx .05$. Thus, x satisfies $(x-125)/25 = 1.64$ which is $x \approx 166$.

Example (3-8 continue)

```
> pnorm(180,mean=125,sd=25)
-pnorm(150,mean=125,sd=25)
[1] 0.1447518
> pnorm(2.2)-pnorm(1)
[1] 0.1447518
> (pnorm(2.2)-pnorm(1))*500
[1] 72.3759
> qnorm(0.95,mean=125,sd=25)
[1] 166.1213
> 25*qnorm(0.95)+125
[1] 166.1213
>
> x<-seq(-5,5,by=0.01)
> plot(x,dnorm(x),type='l',col='red',
xlab='x',ylab='',
main='Standard normal distribution')
> dev.copy2pdf(file='fig3-5.pdf')
```



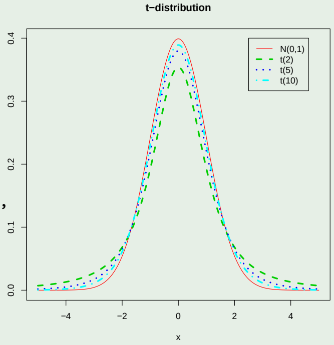
3.10.2. t distribution

t distribution is the probability distribution for a random variable $T = Z/\sqrt{V/r}$ consisting of two independent random variables, $Z \sim \mathrm{N}(0,1)$ and $V \sim \chi^2(r)$.

It is bell-shaped and centered at 0, which is the mean of t distribution, as the standard normal distribution. However t distribution has thicker tails. t distribution gets closer to the standard normal distribution when the degree of freedoms, r, gets larger.

Example (3-9)

```
> x<-seq(-5,5,by=0.01)
> plot(x,dnorm(x),type='l',col='red',
xlab='x',ylab='',main='t-distribution')
> lines(x,dt(x,2),lty=2,col=3,lwd=3)
> lines(x,dt(x,5),lty=3,col=4,lwd=3)
> lines(x,dt(x,10),lty=4,col=5,lwd=3)
> legend(2.5,0.4, c('N(0,1)','t(2)','t(5)',
't(10)'),col=c(2,3,4,5),lty=c(1,2,3,4),
lwd=c(1,3,3,3))
> dev.copy2pdf(file='fig3-6.pdf')
```



3.10.3. Exponential distribution

A random variable X follows **exponential distribution** when its probability density function is

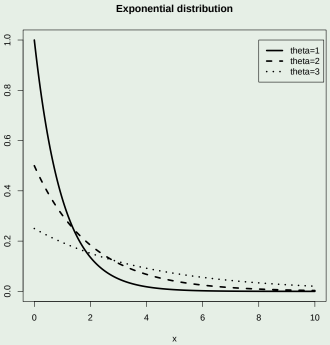
$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0$$

with $\theta > 0$.

It is easily shown that $E(X) = \theta$ and $var(X) = \theta^2$.

Example (3-10)

```
> x<-seq(0,10,by=0.01)
> plot(x,dexp(x,rate=1),type='l',xlab='x',
ylab='',main='Exponential distribution',
lwd=3) # 1/rate=theta=1
> lines(x,dexp(x,rate=0.5),lty=2,lwd=3)
# theta=2
> lines(x,dexp(x,rate=0.25),lty=3,lwd=3)
# theta=4
> legend(8,1,legend=c('theta=1','theta=2',
'theta=3'),lty=1:3,lwd=3)
> dev.copy2pdf(file='fig3-7.pdf')
```



3.10.4. Gamma distribution and chi-square distribution

Gamma distribution has the probability density function

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0$$

with $\alpha > 0$ and $\beta > 0$. We denote $X \sim \text{Gamma}(\alpha, \beta)$.

$\Gamma(\alpha)$ is the normalizing constant whose form is

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

Using $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, the expectation and variance can be easily calculated:

$$E(X) = \alpha\beta, \quad var(X) = \alpha\beta^2.$$

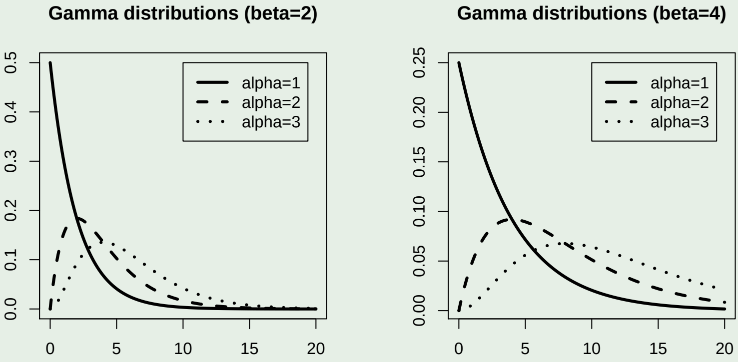
Chi-square distribution is a special case of gamma distributions. A chi-square distribution with the degree of freedoms n, $\chi^2(n)$, is the gamma distribution with $\alpha = n/2$ and $\beta = 2$, $\text{Gamma}(n/2, 2)$. Thus, the expectation and variance of $X \sim \chi^2(n)$ are

$$E(X) = n, \quad var(X) = 2n.$$

The exponential distribution, $\text{Exp}(\theta)$, is also a special case of gamma distribution, $\text{Gamma}(1, \theta)$.

Example (3-11)

```
> x<-seq(0,20,by=0.01)
> par(mfrow=c(1,2))
> plot(x,dgamma(x,shape=1,scale=2),type='l',lwd=3,xlab='',ylab='',main='Gamma distributions (beta=2)')
> lines(x,dgamma(x,shape=2,scale=2),lty=2,lwd=3)
> lines(x,dgamma(x,shape=3,scale=2),lty=3,lwd=3)
> legend(10,0.5,legend=c('alpha=1','alpha=2','alpha=3'),lty=1:3,lwd=3)
> plot(x,dgamma(x,shape=1,scale=4),type='l',lwd=3,xlab='',ylab='',main='Gamma distributions (beta=4)')
> lines(x,dgamma(x,shape=2,scale=4),lty=2,lwd=3)
> lines(x,dgamma(x,shape=3,scale=4),lty=3,lwd=3)
> legend(10,0.25,legend=c('alpha=1','alpha=2','alpha=3'),lty=1:3,lwd=3)
> dev.copy2pdf(file='fig3-8.pdf')
```



3.10.5. Inverse gamma distribution and inverse chi-square distribution

The distribution of $Y = 1/X$ is the **inverse gamma distribution** when $X \sim \text{Gamma}(\alpha, \beta)$. We denote $Y \sim \text{IG}(\alpha, \beta)$. Its probability density function is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{-\alpha-1} e^{1/(\beta x)}, \quad x > 0.$$

Its expectation and variance are

$$E(Y) = \frac{1}{(\alpha-1)\beta} \ (\alpha > 1), \qquad var(Y) = \frac{1}{(\alpha-1)^2(\alpha-2)\beta^2} \ (\alpha > 2).$$

For $X \sim \text{Gamma}(n/2, 2)$, $Y = 1/X$ has the **inverse chi-square distribution**.

3.10.6. Beta distribution

If a continuous random variable X, defined on the unit interval $[0,1]$, has the probability density function

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1,$$

then X has **beta distribution** and we denote $X \sim \mathrm{B}(\alpha,\beta)$.

Here, $B(\alpha,\beta)$ is the **beta function** that

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

If $\alpha = \beta$, then the distribution is symmetric about $1/2$. If $\alpha < \beta$ then it is right-skewed. If $\alpha > \beta$ then it is left-skewed.

The expectation and variance for $\mathrm{B}(\alpha,\beta)$ are

$$E(X) = \frac{\alpha}{\alpha+\beta}, \quad var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Example (3-12)

```
> x<-seq(0,1,by=0.01)
> par(mfrow=c(1,3))
> plot(x,dbeta(x,3,3),type='l',xlab='x',ylab='',main='Beta distribution: alpha=beta')
> plot(x,dbeta(x,4,2),type='l',xlab='x',ylab='',main='Beta distribution: alpha>beta')
> plot(x,dbeta(x,2,4),type='l',xlab='x',ylab='',main='Beta distribution: alpha<beta')
> dev.copy2pdf(file='fig3-9.pdf')
```