Special Random Variables

Mahbub Latif, PhD

January 2024

Plan

- The binomial distribution

- The Poisson distribution

- Hypergeometric distribution

The Binomial Random Variable

The Bernoulli Random Variable

- Suppose that a trial has only two outcomes, which are classified as either a "success" or as a "failure", such trial is known as **Bernoulli trial**

  ○ Define a random variable X as

$$X = \begin{cases} 1 & \textit{if the outcome is a success} \\ 0 & \textit{if the outcome is a failure} \end{cases}$$

The Bernoulli Random Variable

- X follows a Bernoulli distribution with parameter p and the probability mass function

$$P(X = 1) = p$$
$$P(X = 0) = 1 - p$$

- The probability mass function

$$P(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1; \ \ 0 \le p \le 1$$

  - $E(X) = p$ and $Var(X) = p(1 - p)$

The binomial random variable

- Binomial random variable deals with the distribution of the number of successes in n independent Bernoulli trials

- The probability success p remain constant from trial to trial

- Let $X_1, \ldots, X_n$ are independent and each follow a Bernoulli distribution with parameter p, and define

$$X = X_1 + \cdots + X_n$$

- X represents the number of success in n Bernoulli trials and X follows a binomial distribution with parameters n and p, i.e. $X \sim B(n, p)$

The binomial random variable

- The probability mass function of X

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

  - $\binom{n}{x} \rightarrow$ number of ways x success can be obtained in n Bernoulli trials

Example 5.1a

- It is known that disks produced by a certain company will be defective with probability .01 independently of each other.

- The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective.

- What proportion of packages is returned?

- If someone buys three packages, what is the probability that exactly one of them will be returned?
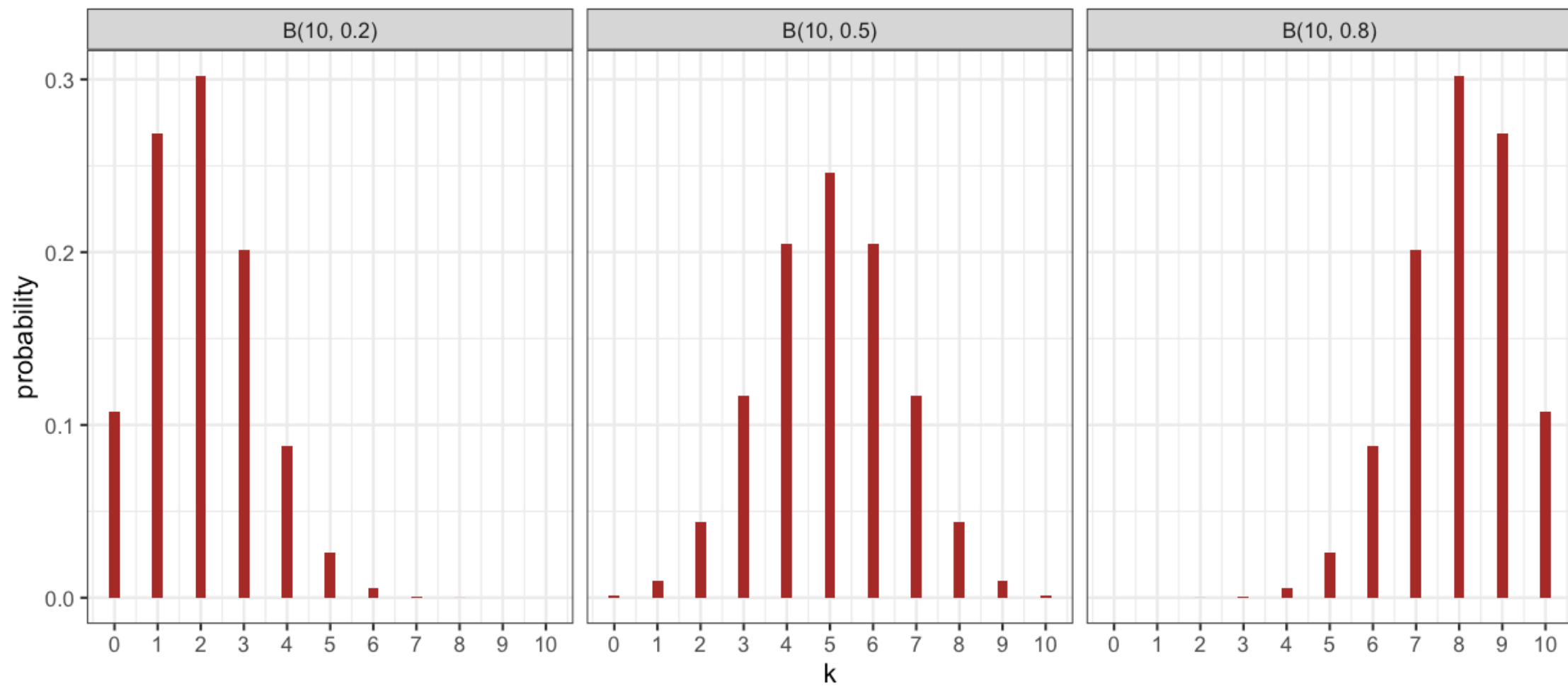
Expectation of binomial distribution

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^k (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np$$

Variance of binomial distribution

$$E\big[X(X-1)\big] = \sum_{x=0}^{n} x(x-1)\binom{n}{x}p^k(1-p)^{n-x}$$

$$= \sum_{x=2}^{n} x(x-1)\binom{n}{x}p^x(1-p)^{n-x} = n(n-1)p^2$$

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= E\big[X(X-1)\big] + E(X) - [E(X)]^2$$
$$= n(n-1)p^2 + np - n^2p^2$$
$$= np(1-p)$$

Binomial cumulative distribution function

- The cumulative distribution function of $X \sim B(n, p)$

$$P(X \le a) = \sum_{x=a}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

- Following relationship is helpful to calculate cumulative distribution of binomial distribution

$$P(X = x + 1) = \frac{p}{1-p} \frac{n-x}{x+1} P(X = x)$$

Problems

- 1, 3, 5, 6, 7, 9
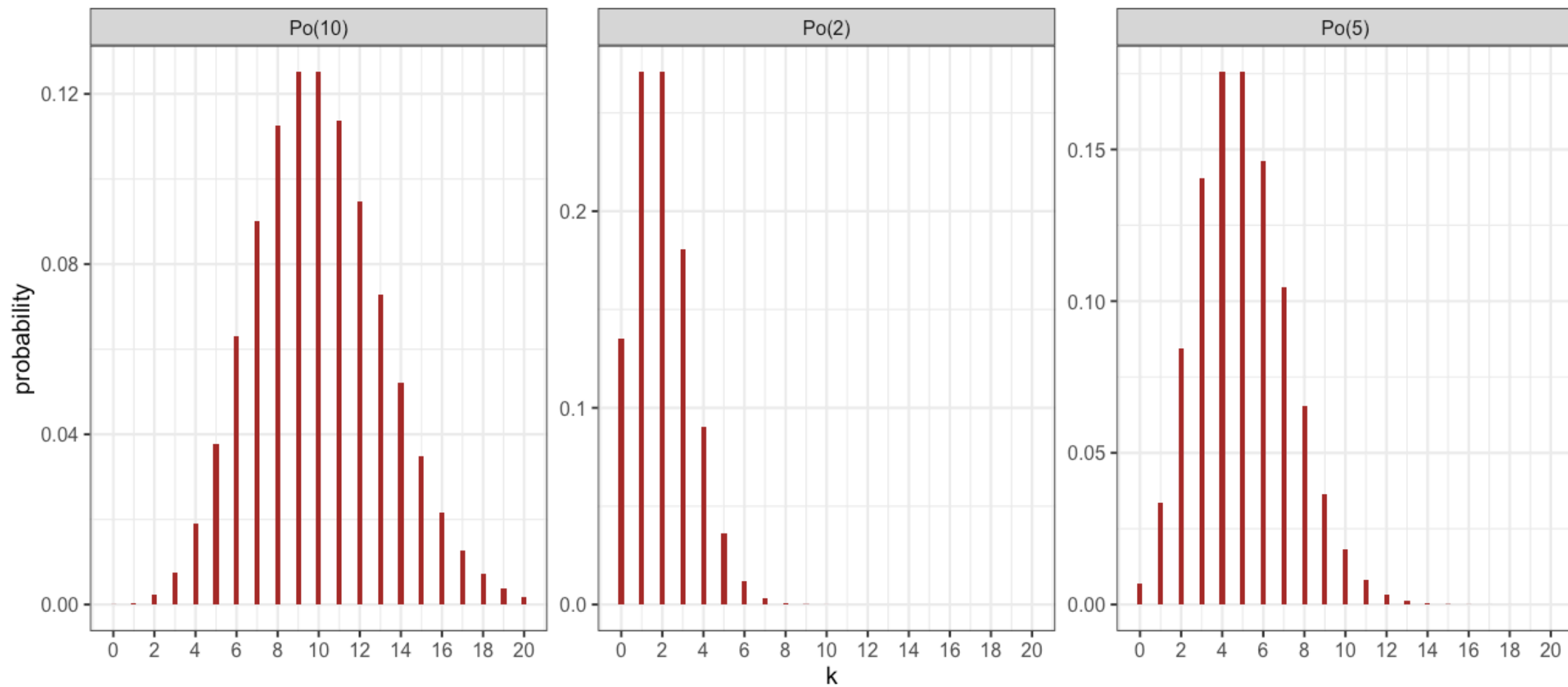
The Poisson Random Variable

The Poisson Random Variable

- A random variable X is said to be a Poisson random variable with parameter $\lambda > 0$, if its probability mass function is given by

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

- Using the relationship $e^\lambda = \sum_{x=0}^{\infty} \lambda^x/x!$, it can be shown that $\sum_{x=0}^{\infty} P(X = x) = 1$ for a Poisson distribution

- For $X \sim Po(\lambda)$

$$E(X) = V(X) = \lambda$$

Examples of Poisson random variables

- The number of misprints on a page of a book

- The number of people in a community living to 100 year of age

- The number of wrong telephone numbers that are dialed in a day

- The number of transistors that fail on their first day of use

- The number of customers entering a post office on a given day

Expectation and variance of a Poisson random variable

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!} = \lambda$$

Similarly

$$E\big[X(X-1)\big] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2$$

$$Var(X) = E\big[X(X-1)\big] + E(X) - [E(X)]^2 = \lambda$$

Poisson approximation of binomial distribution

- Suppose $X \sim B(n, p)$, for a large n and small p

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \simeq \frac{e^{-\lambda} \lambda^x}{x!}$$

  - $\lambda = np$

- **Examples:** $X \sim B(70, .1)$

$$P(X = 5) = \begin{cases} 0.1284 & \textit{for binomial} \\ 0.1277 & \textit{for Poisson} \end{cases}$$

Example 5.2a

- Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3.

- Calculate the probability that there is at least one accident this week.

Example 5.2b

- Suppose the probability that an item produced by a certain machine will be defective is .1.

- Find the probability that a sample of 10 items will contain at most one defective item.

- Assume that the quality of successive items is independent.

Example 5.2d

- If the average number of claims handled daily by an insurance company is 5, what proportion of days have less than 3 claims?

- What is the probability that there will be 4 claims in exactly 3 of the next 5 days?

- Assume that the number of claims on different days is independent.

Distribution of sum of two independent Poisson random variables

- If $X_1 \sim Po(\lambda_1)$ and $X_2 \sim Po(\lambda_2)$, then

$$Y = X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$$

- Moment generating function of X_1

$$M_{X_1}(t) = E\left[e^{tX}\right] = \exp\left[\lambda_1(e^t - 1)\right]$$

- Moment generating function of $Y = X_1 + X_2$

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t) = \exp\left[(\lambda_1 + \lambda_2)(e^t - 1)\right]$$

Example 5.2f

- It has been established that the number of defective stereos produced daily at a certain plant is Poisson distributed with mean 4.

- Over a 2-day span, what is the probability that the number of defective stereos does not exceed 3?

Problems

- 13, 14, 18

Hypergeometric Distribution

Hypergeometric Distribution

- A bin contains $N + M$ batteries, of which N are of acceptable quality and M are defective

- A sample of size n is randomly chosen (without replacement) and all possible sampled subsets are equally likely

- The X denotes the number of acceptable batteries in the sample of size n and its probability mass function

$$P(X = x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{N+M}{n}}, \ k = 0, 1, \ldots, \min(n, N)$$

- X follows a hypergeometric distribution with parameters N, M, and n

Example 5.3a

- The components of a 6-component system are to be randomly chosen from a bin of 20 used components.

- The resulting system will be functional if at least 4 of its 6 components are in working condition.

- If 15 of the 20 components in the bin are in working condition, what is the probability that the resulting system will be functional?

- X, the number of components with working condition out of six selected components, follows a hypergeometric distribution with parameters $N = 15$, $M = 5$, and $n = 6$

Hypergeometric Distribution

- Let $X \sim B(n, p)$ and $Y \sim B(m, p)$ then

$$P\Big(X = i \mid X + Y = k \Big) = \frac{\binom{n}{i}\binom{m}{k-i}}{\binom{n+m}{k}}$$

The Uniform Random Variable

The Uniform Random Variable

- A random variable X is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its pdf is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

- Show that

$$E(X) = \frac{\alpha + \beta}{2} \quad and \quad Var(X) = \frac{(\beta - \alpha)^2}{12}$$

The Normal Random Variable

The Normal Random Variable

- A random variable X is said to be normally distributed random variable with parameters μ and σ^2,

  ○ i.e. $X \sim N(\mu, \sigma^2)$, if its density is

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

- The normal density function is bell-shaped and symmetric about its mean μ

- The maximum value of the density function is $(\sigma\sqrt{2\pi})^{-1} \simeq 0.399/\sigma$ attains at $x = \mu$

Parameters of normal distribution and its density function
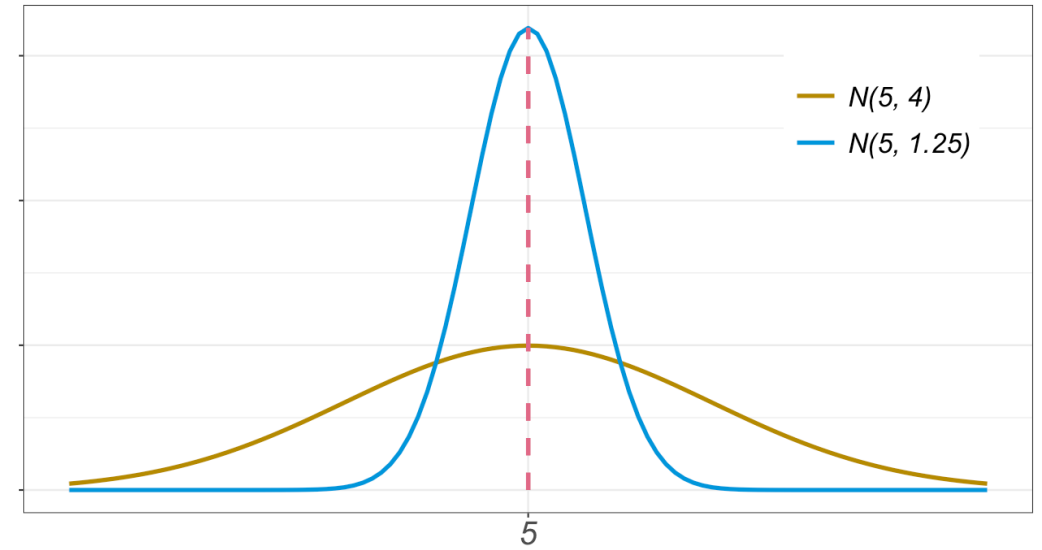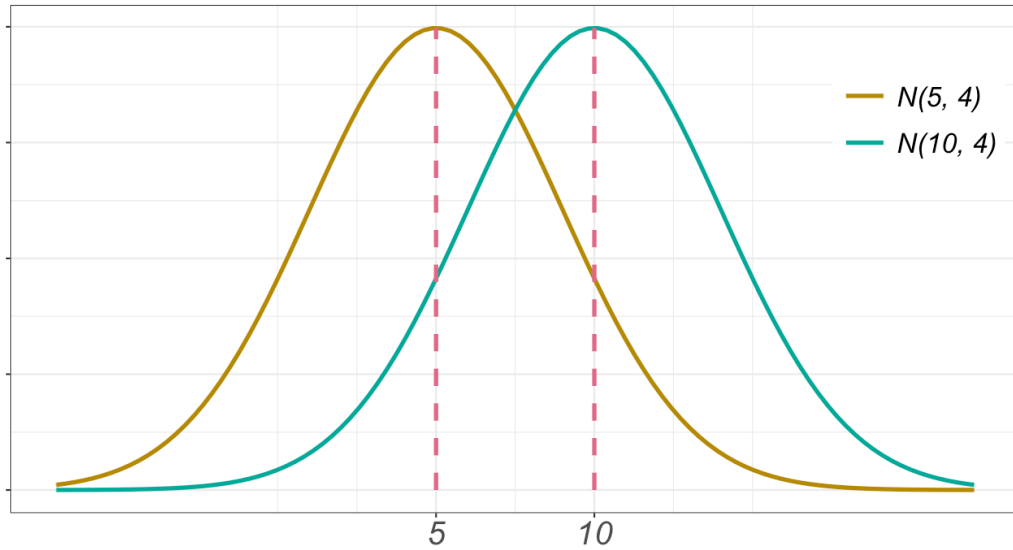
The normal random variable

- The expected value

$$E(X - \mu) = \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = 0$$

$$E(X) = \mu$$

- The variance

$$Var(X) = E(X - \mu)^2$$

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

Distribution of a linear combination of a normal distribution

- If $X \sim N(\mu, \sigma^2)$, what is the distribution of $Y = a + bX$, where a and b are constants?

- If $M(t)$ is the moment generating function (mgf) of X, the mgf of $Y = a + bX$

$$M_Y(t) = \int e^{ty} f(x) dx = \int e^{t(a+bx)} f(x)\, dx = e^{ta} M_X(tb)$$

Distribution of a linear combination of a normal distribution

- The mgf of $X \sim N(\mu, \sigma^2)$

$$M_X(t) = e^{t\mu + (t^2\sigma^2/2)}$$

- The mgf of $Y = a + bX$

$$M_Y(t) = e^{ta} M_X(tb)$$
$$= e^{ta} e^{t\mu b + (t^2 b^2 \sigma^2/2)}$$
$$= e^{t(b\mu + a) + (t^2 b^2 \sigma^2/2)}$$

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

Standard normal distribution

- A normal distribution with mean 0 and variance 1 is known as the standard normal distribution, which is denoted by Z

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

- The mgf of standard normal distribution: $M_Z(t) = e^{t^2/2}$

- The probability density function of a standard normal distribution

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

Standard normal distribution

- If $X \sim N(\mu, \sigma^2)$ then

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

  ○ Proof this relationship!

- It can be shown that

$$E(Z) = E\left[\frac{X - \mu}{\sigma}\right] = 0 \ \ and \ \ Var(Z) = Var\left[\frac{X - \mu}{\sigma}\right] = 1$$

Cumulative distribution of the standard normal distribution

- The cdf of the standard normal distribution

$$\Phi(x) = P(Z < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

  - This integration cannot be evaluated agebraically

  - Most of the matematical statistics book has Z table, which provides $\Phi(x)$ values for different x (see page 642 of the textbook for the Z table)

- An important relationship

$$\Phi(x) + \Phi(-x) = 1$$

Probability calculation related to $Z \sim N(0, 1)$

$$P(Z < 1.12) = \Phi(1.12)$$
$$= 0.8686$$
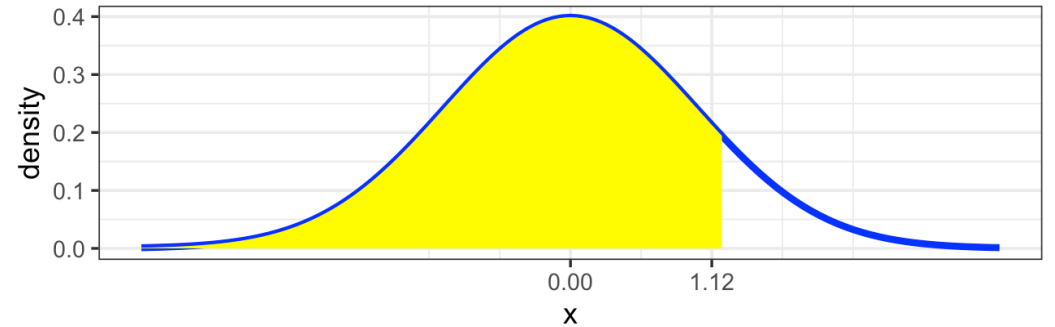
$$P(Z > .73) = 1 - P(Z \leq .73)$$
$$= 1 - \Phi(.73)$$
$$= 1 - 0.7673$$
$$= 0.2327$$
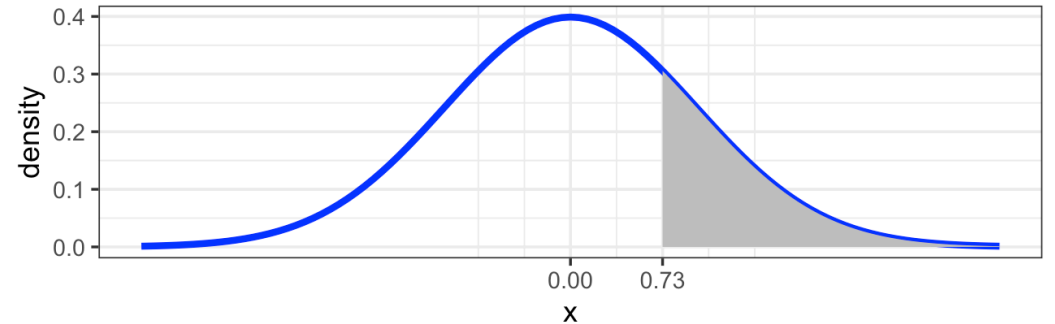
Probability calculation related to $Z \sim N(0, 1)$

$$P(Z < -1.22) = \Phi(-1.22)$$
$$= 1 - \Phi(1.22)$$
$$= 1 - 0.8686$$
$$= 0.1112$$
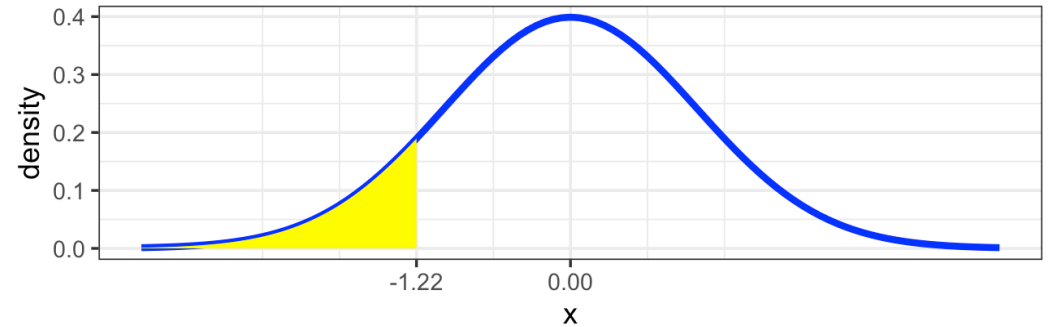
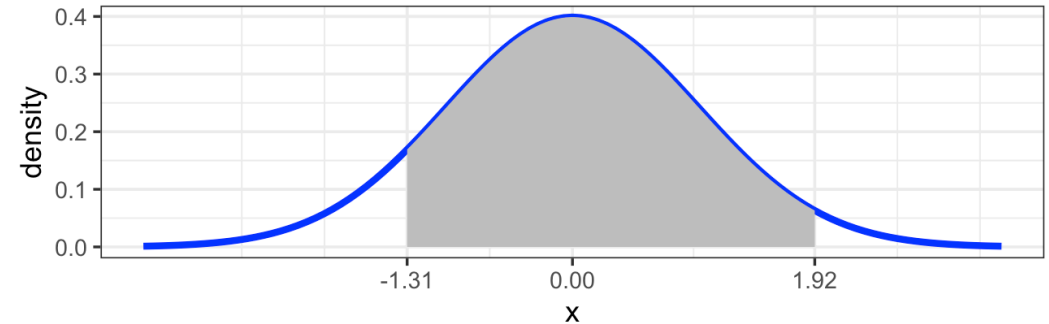$$P(Z > -1.13) = 1 - P(Z \leq -1.13)$$
$$= 1 - \Phi(-1.13)$$
$$= \Phi(1.13)$$
$$= 0.8708$$
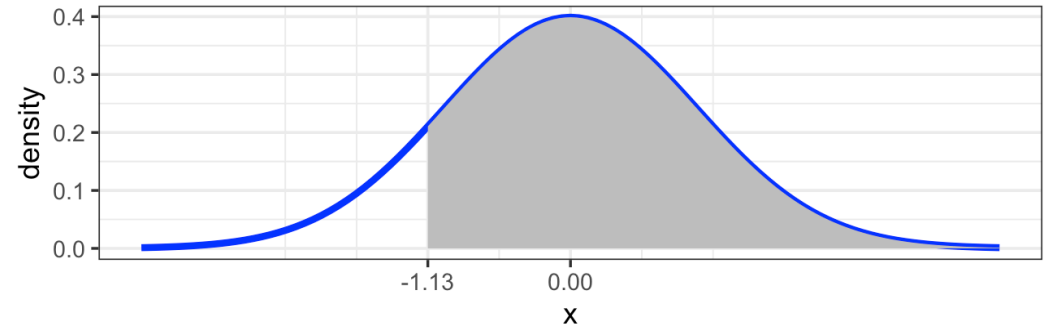
Probability calculation related to $Z \sim N(0, 1)$

$$P(1.1 < Z < 2.2) = \Phi(2.2) - \Phi(1.1)$$
$$= 0.9861 - 0.8643$$
$$= 0.1218$$
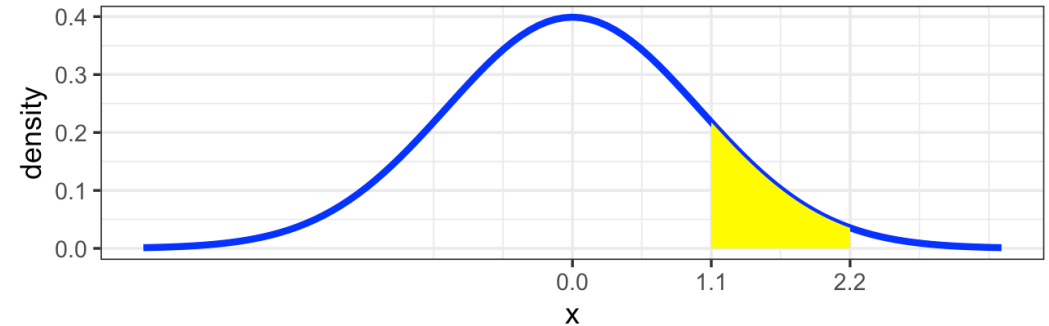
$$P(-1.31 < Z < 1.92)$$
$$= \Phi(1.92) - \Phi(-1.31)$$
$$= \Phi(1.92) - 1 + \Phi(1.31)$$
$$= 0.9726 - 1 + 0.9049$$
$$= 0.8775$$

Probability calculation related to $X \sim N(\mu, \sigma^2)$

- For any $a < b$

$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma} \right)$$

$$= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma} \right)$$

$$= \Phi\left(\frac{b - \mu}{\sigma} \right) - \Phi\left(\frac{a - \mu}{\sigma} \right)$$

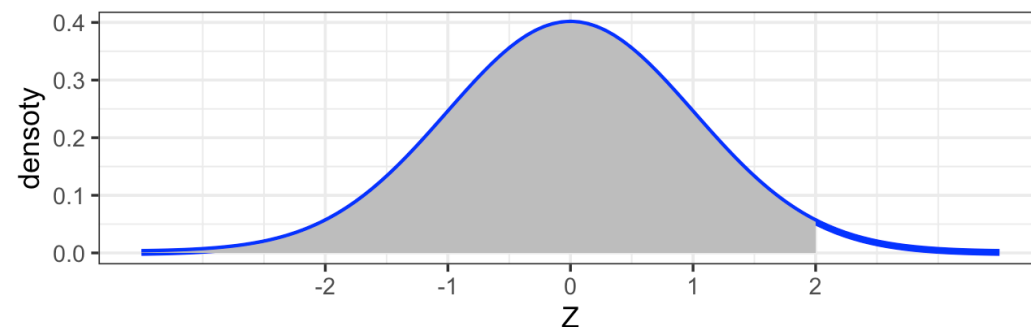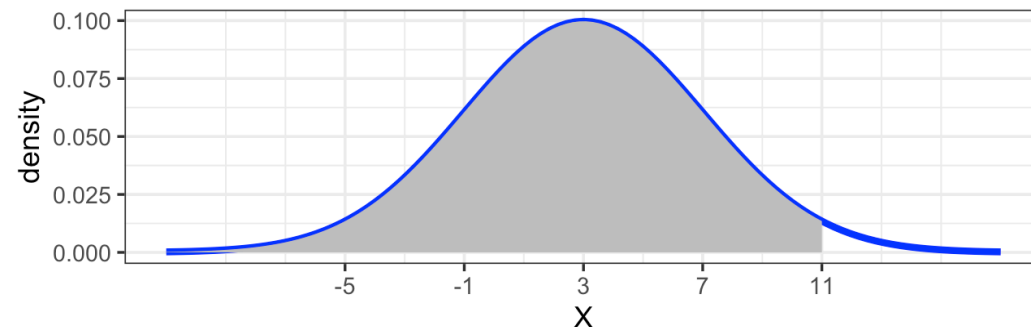# Probability calculation related to $X \sim N(\mu, \sigma^2)$

- For $X \sim N(3, 16)$

$$P(X < 11) = \Phi\left(\frac{11 - 3}{4}\right)$$

$$= \Phi(2)$$

$$= 0.9772$$

- Calculate

$$P(X > -1)$$

$$P(2 < X < 7)$$

Sum of several independent normal variables

- Let $X_1, \ldots, X_n$ be n independent normal random variables, where

$$X_i \sim N(\mu_i, \sigma_i^2)$$

- The distribution of $Y = \sum_{i=1}^{n} X_i$ follows a normal distribution with

$$mean = \sum_{i=1}^{n} \mu_i \quad and \quad variance = \sum_{i=1}^{n} \sigma_i^2$$

Exammple 5.5d

- Data from National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angles is a normal random variable with a mean of 12.08 inches and standard deviation of 3.1 inches.

- Find the probability that total precipitation during the next two years will exceed 25 inches.

Quantiles/percentiles of standard normal distribution

- $z_\alpha \to$ the $100(1 - \alpha)$ percentile of standard normal distribution, i.e.,

$$P(Z < z_\alpha) = 1 - \alpha$$

- From the table, we can find

  ○ $z_{.05} = 1.645$

  ○ $z_{.025} = 1.96$

  ○ $z_{.01} = 2.33$
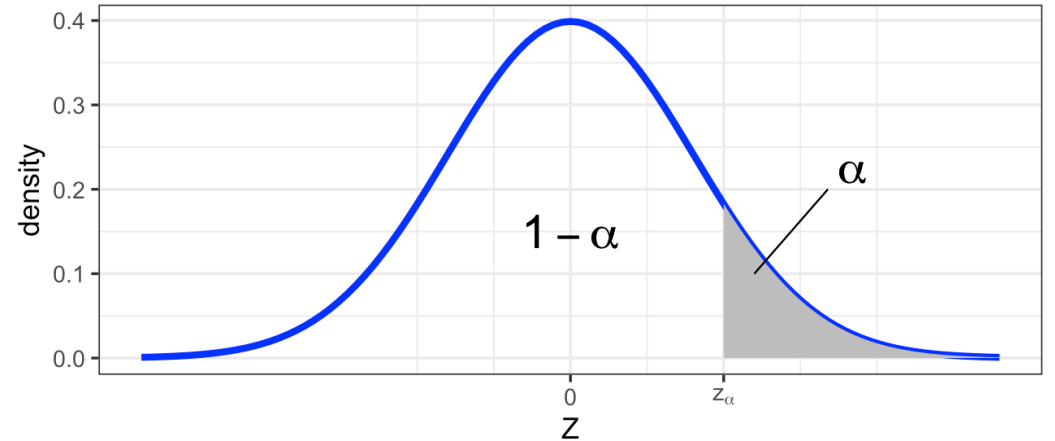
Exponential random variables

Exponential random variables

- An exponential random variable with parameter $\lambda > 0$ has the following probability density function
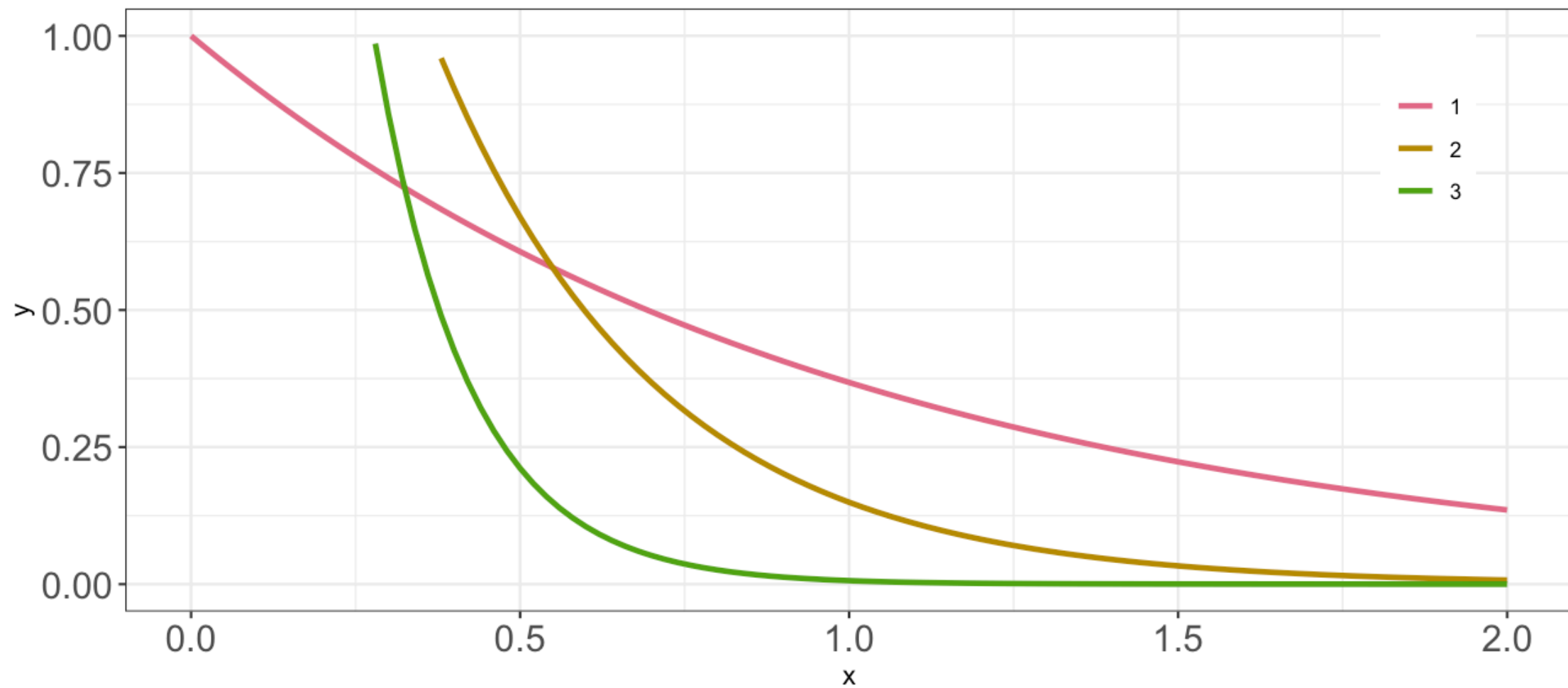
$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

- The moment generating function of a exponential random variable

$$M_X(t) = \lambda/(\lambda - t)$$

- Show that

$$E(X) = 1/\lambda \ \ and \ \ Var(X) = 1/\lambda^2$$

The cumulative distribution function

$$F(x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}$$

- $P(a < X < b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$

Memoryless property of exponential random variable

- If X follows an exponential distribution with parameter λ

$$P(X > s + t \,|\, X > t) = P(X > s)$$

Problems

- 1, 3, 5, 6, 7, 8, 10, 11, 13, 14, 23, 24, 25, 26, 33, 34, 35, 26, 37, 38