Chapter 4 R Practice

Commands for continuous random variables are similar to those for discrete random variables. As noted in Lab 05, the abbreviated name of the distribution is preceded by once of four letters:

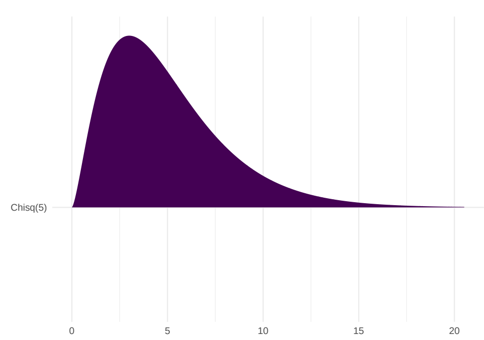
- "d" for the density function (pdf for continuous random variables)
- "p" for the distribution function (cdf)
- "q" for quantile function (gives the quantile for a given probability)
- "r" for random (generates a random number based on the distribution

Computing probabilities for continuous distributions in R

The cdf values $F(Y) = P(Y < y)$ are returned with commands using "p" at the beginning of the R distribution name:

- `punif(y, min, max)` returns $F(y)$ values for a **uniform random variable** $Y \sim U(a, b)$ where the arguments `min` and `max` correspond to the parameters a and b respectively.
- `pexp(y, rate)` - returns $F(y)$ values for an **exponential random variable** $Y \sim exp(\theta)$ with argument `rate` $= \lambda = 1/\theta$
  - Note our textbook defines exponential distributions in terms of the parameter θ, which is the mean wait time, instead of the rate λ.
- `pgamma(y, shape, scale = 1/rate)` - returns $F(y)$ values for a **gamma random variable** $Y \sim (\alpha, \theta)$ with parameters `shape` $= \alpha$ and `scale` $= \theta = 1/\lambda$.
  - Because the gamma function can be parameterized in terms of shape or rate, make sure to name the argument inside the `pgamma()` function to specify which parameter you're defining.
- `pchisq(y, df)` - returns $F(y)$ values for a **chi-square random variable** $Y \sim \chi^2(r)$ where the argument `df` corresponds to the degrees of freedom parameter `r`
- `pnorm(y, mean, sd)` - returns $F(y)$ values for a **normal random variable** $Y \sim N(\mu, \sigma^2)$ where the argument `mean` corresponds to the parameter μ and the argument `sd` refers to the standard deviation $\sqrt{\sigma^2}$.

There is a hidden argument in each of these functions called `lower.tail`, and the default is set to `lower.tail = TRUE`. Meaning, by default, the "p" functions return the probabilities in the *lower tail* of the density function: $F(Y) = P(Y < y)$. Sometimes it will be useful to override the default by adding the argument `lower.tail = FALSE`, which will return values $P(Y > y)$. You can think of these two options visually; `lower.tail = TRUE` represents the dark shaded region $P(Y < 10)$, and `lower.tail = FALSE` represents the light shaded region $P(Y > 10)$ in the plot below.



Uniform example

Suppose $Y \sim U(0, 5)$.

1. Sketch the pdf and shade the region $Y < 2$

2. Write the code you need to calculate the probability that $Y < 2$. *Hint: use the* `punif()` *function.* How many/what arguments are you required to specify?

3. Calculate $P(Y < 2)$ (by hand or using R)

4. Write out two different lines of code you could use to find $P(Y > 2)$ in R. *Hint: one should make use of the* `lower.tail` *argument.* Execute both lines of code in R to verify that they give the value you expect.

5. Suppose we want to find $P(2 < Y < 3)$. Sketch a graph with this region shaded.

6. Using your visual as a guide, what would you need to subtract from $P(Y < 3)$ in order to obtain $P(2 < Y < 3)$?

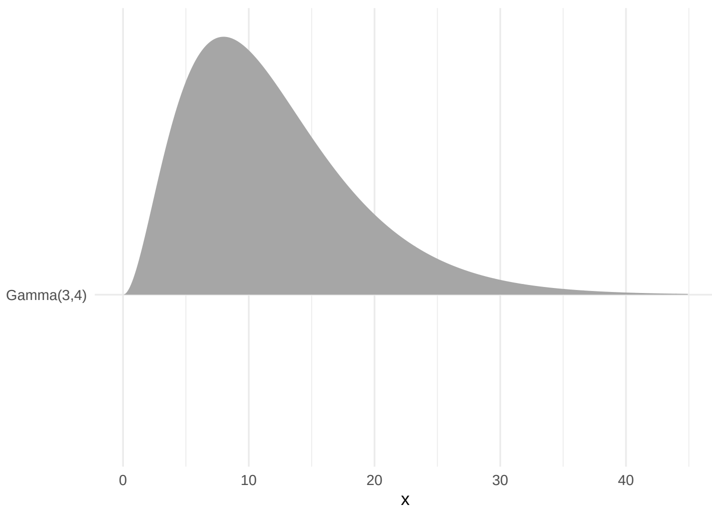
7. Using your answer to part 6 as a guide, write out R code for computing $P(2 < Y < 3)$. Execute your code in R and report the answer.

Gamma example

Suppose $Y \sim \Gamma(\alpha = 3, \theta = 4)$.

We can plot this distribution with the following code:

```
ggplot() +
  stat_dist_slab(aes(y = "Gamma(3,4)",
                     #note the function dist_gamma() requires the rate parameter
                     #instead of the scale parameter. Recall rate = 1/scale
                     dist = dist_gamma(shape = 3, rate = 1/4))) +
  labs(y = "", y = "y") +
  theme_minimal()
```



8. What's the mean of this distribution? Does this seem reasonable based on the graph above?

9. Shade the region $Y < 5$ on the graph above

10. Write out the code you need to calculate $P(Y < 5)$. *Hint: use the `pgamma()` function.* How many/what arguments do you need to specify?

11. Use your code above to compute $P(Y < 5)$ and report the answer. Does the value seem reasonable based on the plot above?

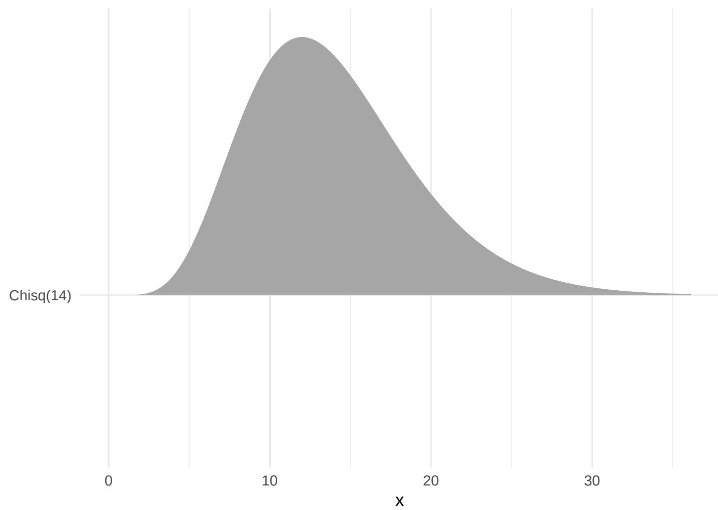
12. Write out two different lines of code you could use to find $P(Y > 5)$ in R. *Hint: one should make use of the `lower.tail` argument.* Execute both lines of code in R to verify they give the value you expect.

Chi-squared example

Suppose $Y \sim \chi^2(14)$.

We can plot this distribution with the following code:

```
ggplot() +
  stat_dist_slab(aes(y = "Chisq(14)",
                     dist = dist_chisq(14))) +
  labs(y = "", y = "y") +
  theme_minimal()
```



13. What's the mean of this distribution? Does this seem reasonable based on the graph above?

14. Shade the region $Y < 10$ on the graph above

15. Write out the code you need to calculate $P(Y < 10)$. *Hint: use the* `pchisq()` *function.* How many/what arguments do you need to specify?

16. Use your code above to compute $P(Y < 10)$ and report the answer. Does the value seem reasonable based on the plot above?

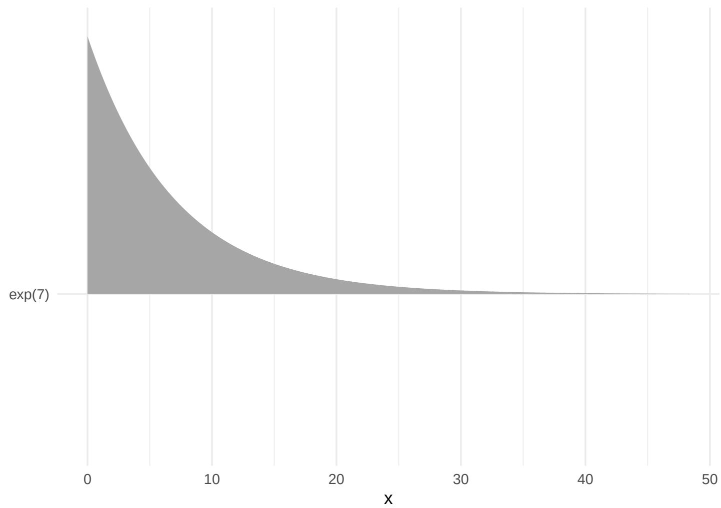
17. Write out two different lines of code you could use to find $P(Y > 10)$ in R. *Hint: one should make use of the* `lower.tail` *argument.* Execute both lines of code in R to verify they give the value you expect.

Exponential example

Suppose $Y \sim exp(7)$.

We can plot this distribution with the following code:

```
ggplot() +
  stat_dist_slab(aes(y = "exp(7)",
                              #note the function dist_exponential() requires the rate parameter
                              #instead of the scale parameter. Recall rate = 1/scale
                              dist = dist_exponential(1/7))) +
  labs(y = "", y = "y") +
  theme_minimal()
```



18. What's the mean of this distribution? Does this seem reasonable based on the graph above?

19. Write out the code you need to calculate $P(Y > 20)$. *Hint: use the* `pexp()` *function.* How many/what arguments do you need to specify?

20. Use your code above to compute $P(Y > 20)$ and report the answer. Does the value seem reasonable based on the plot above?

21. Shade $P(10 < Y < 20)$ on the plot above.

22. Write out the code you need to find $P(10 < Y < 20)$. Execute the code and report the answer. Does it seem reasonable based on the plot above?

Computing percentiles for continuous distributions in R

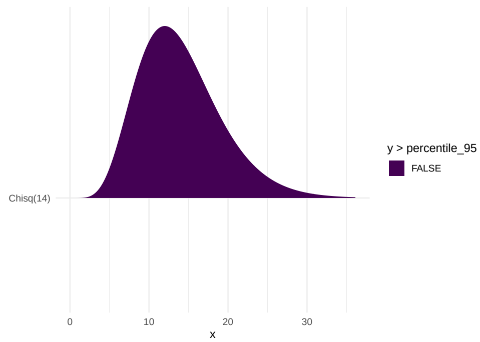
The inverse of the cdf, or the y that solves the equation $p = F(y) = P(Y < y)$, are returned with commands using "q" at the beginning of the R distribution name:

- `qunif(p, min, max)` – finds y such that $P(Y < y) = p$ for a uniform random variable $Y \sim U(a, b)$ where the arguments `min` and `max` correspond to the parameters a and b respectively.
- `qexp(p, rate)` - finds y such that $P(Y < y) = p$ for an exponential random variable $Y \sim exp(\lambda)$ with argument `rate` $= \lambda = 1/\theta$
- `qgamma(p, shape, scale = 1/rate)` - finds y such that $P(Y < y) = p$ for a gamma random variable $Y \sim (\alpha, \theta)$ with parameters `shape` $= \alpha$ and `scale` $= \theta = 1/\lambda$.
- `qchisq(p, df)` - finds y such that $P(Y < y) = p$ for a chi-square random variable $Y \sim \chi^2(r)$ where the argument `df` corresponds to the degrees of freedom parameter `r`.

Again, set `lower.tail = FALSE` to use upper tail probabilities and find y such that $P(Y > y) = p$.

Chi-squared example

Suppose $Y \sim \chi^2(14)$ as in the example above. Below is a visual representation of the 95th percentile.



23. Write out the code you need to find the 95th percentile in R. Execute the code and report the value. Does the value seem reasonable based on the plot above?

Gamma example

Suppose $Y \sim \Gamma^2(\alpha = 3, \theta = 7)$. Below is a visual of the distribution shaded at the median.

24. Write out the code you need to find the median in R. Execute the code and report the value. Does the value seem reasonable based on the plot above?

Normal Distribution Practice

25. Assume $Z \sim N(0, 1)$. Find the following:

a. $P(Z \leq 1.25)$

b. $P(1.25 \leq Z \leq 2.31)$

c. $P(-2.31 \leq Z \leq -1.25)$

d. $P(Z \leq -2.31)$

e. $P(-2.31 \leq Z \leq 1.25)$

f. $P(Z > 1.25)$

g. If you haven't already, draw pictures of each of the probabilities you computed above. How do b & c relate to one another?

h. What about a, d, & e?

i. What about a & f?

26. The textbook defines a quantity z_α as: $P(Z \geq z_\alpha) = \alpha$. (This is unusual as it is the $(1-\alpha)th$ percentile). Find:

a. $z_{0.9147}$

b. $z_{0.0125}$

c. $z_{0.05}$

d. $-z_{0.025}$

27. If Y is normally distributed with a mean of 6 and a variance of 25, find:

a. $P(6 \leq Y \leq 14)$

b. $P(4 \leq Y \leq 14)$

c. $P(-4 \leq Y \leq 0)$

d. $P(Y > 15)$

e. $P(|Y-6| < 5)$

f. $P(|Y-6| < 10)$