Parametric Distributions

PSE 3.1

In some cases, a researcher may know (or be willing to assume that they know) the distribution of a random variable. More commonly, a researcher might be willing to assume that they know the distribution of a random variable up to a finite number of **parameters**, where the parameters allow for some flexibility in the shape of the distribution. In this section, we will briefly cover some of the most common/useful parametric distributions (though there are a number of other distributions mentioned in this chapter in the textbook that are worth mentioning).

Bernoulli Distribution

PSE 3.2

A random variable that follows a **Bernoulli distribution** takes either the value 0 or 1 with some probability p. An example is whether or not a particular person is employed. In particular, we can write

$$\begin{aligned} \mathrm{P}(X = 1) &= p \\ \mathrm{P}(X = 0) &= 1 - p \end{aligned}$$

which fully summarizes the distribution of X. In practice, in some cases p might be known (e.g., for flipping a coin, $p = 0.5$), but in most interesting cases you would need to somehow estimate p (we'll return to this sort of issue later). It is also useful to have a single, full expression for the pmf of a Bernoulli random variable, which is given by

$$\mathrm{P}(X = x) = p^x (1 - p)^{(1-x)}$$

In particular, just try plugging in $x = 1$ and $x = 0$ here and you will get $\mathrm{P}(X = 1) = p$ and $\mathrm{P}(X = 0) = (1 - p)$ as above.

Two interesting properties of Bernoulli random variables are:

1. $\mathbb{E}[X] = p$, in words: the expected value of X equals p
2. $\mathrm{var}(X) = p(1 - p)$ in words: the variance of X equals $p(1 - p)$.

Here is a proof of the first property, and I'll leave proving the second property as an excerise.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x \in \{0,1\}} x \mathrm{P}(X = x) \\ &= 0(1 - p) + 1p \\ &= p \end{aligned}$$

Binomial Distribution

A **Binomial** random variable counts the number of successes in n independent trials, each of which succeeds with the same probability p. Formally, if $X_1, \dots, X_n$ are iid Bernoulli(p) random variables (e.g., $X_i = 1$ if trial i is a "success" and 0 otherwise), then $X = X_1 + \cdots + X_n$ follows a **Binomial** distribution, written $X \sim \text{Binomial}(n, p)$. For example, if you survey n people and record whether or not each one is employed, the total number of employed people in the sample is Binomial(n, p), where p is the (population) employment rate.

X is discrete with support $\{0, 1, \dots, n\}$, and its pmf is

$$\text{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \dots, n$$

where $\binom{n}{x}$ is the number of ways to choose which x of the n trials were successes (recall combinations from the Counting section of the first set of notes).

```
library(ggplot2)
```

```
x_vals <- 0:10
binom_df <- data.frame(
  x = rep(x_vals, times = 3),
  p = factor(rep(c(0.2, 0.5, 0.8), each = length(x_vals))),
  prob = c(dbinom(x_vals, size = 10, prob = 0.2),
           dbinom(x_vals, size = 10, prob = 0.5),
           dbinom(x_vals, size = 10, prob = 0.8))
)

ggplot(data = binom_df, aes(x = x, y = prob, fill = p)) +
  geom_col(position = "dodge") +
  xlab("x") +
  ylab("pmf") +
  theme_bw()
```

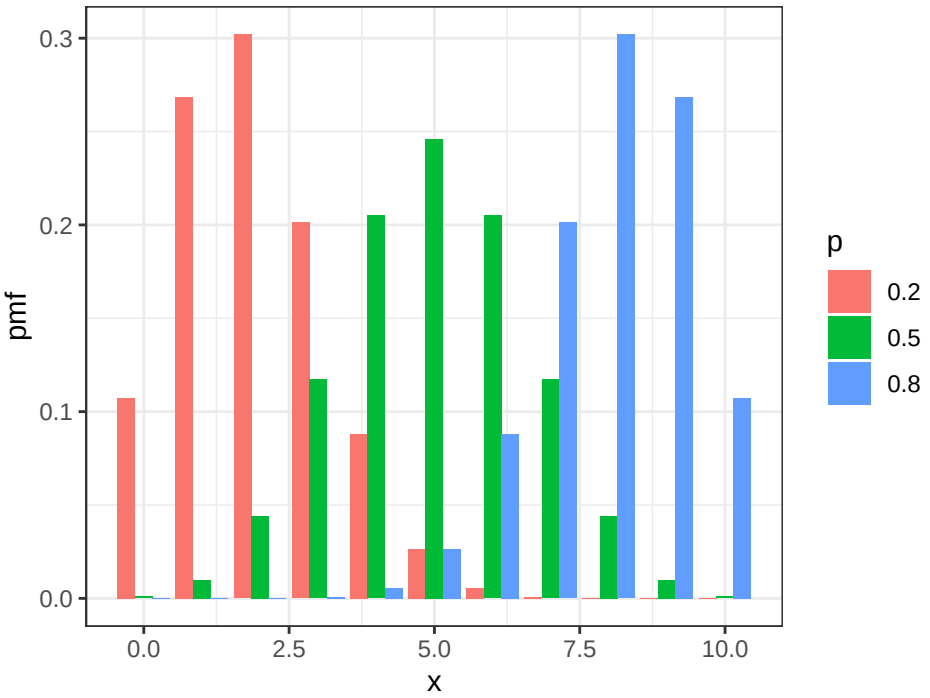

Figure 1: pmf of Binomial(10,p) for a few values of p

Because X is just a sum of n iid Bernoulli(p) random variables, we get $\mathbb{E}[X]$ for free from linearity of expectation (no new calculation needed):

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np$$

It also turns out that $\text{var}(X) = np(1-p)$, but deriving that requires knowing the variance of a sum of independent random variables, which we'll cover in the next set of notes — we'll come back to it there.

In R, `dbinom`/`pbinom`/`qbinom`/`rbinom` give the pmf, cdf, quantile function, and random draws, respectively.

Poisson Distribution

A **Poisson** random variable counts the number of times some event occurs over a fixed interval of time or space, when those events occur independently at a constant average rate $\lambda > 0$. Examples include the number of patents a firm files in a given year, the number of customers arriving at a store in an hour, or the number of workplace accidents at a plant in a given month. We write $X \sim \text{Poisson}(\lambda)$.

X is discrete with support $\{0, 1, 2, \dots\}$ (unlike the Binomial, there is no upper bound), and its

pmf is

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

```
x_vals <- 0:20
pois_df <- data.frame(
  x = rep(x_vals, times = 3),
  lambda = factor(rep(c(1, 4, 10), each = length(x_vals))),
  prob = c(dpois(x_vals, lambda = 1),
           dpois(x_vals, lambda = 4),
           dpois(x_vals, lambda = 10))
)

ggplot(data = pois_df, aes(x = x, y = prob, color = lambda)) +
  geom_line() +
  geom_point() +
  xlab("x") +
  ylab("pmf") +
  theme_bw()
```

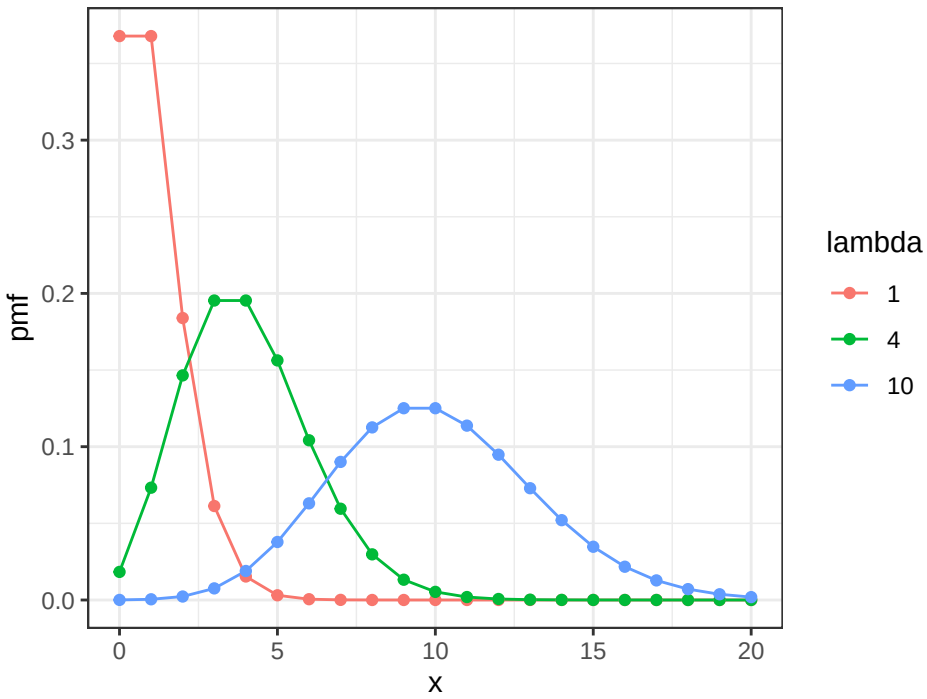

Figure 2: pmf of Poisson(lambda) for a few values of lambda

It turns out that $\mathbb{E}[X] = \lambda$ and $\mathrm{var}(X) = \lambda$ (interestingly, the mean and variance are the same

number for a Poisson random variable). Deriving $\mathbb{E}[X]$ isn't conceptually hard, but it takes a bit of algebra — you re-index the sum and recognize what's left over as the Taylor series $e^{\lambda} = \sum_{x=0}^{\infty} \lambda^x / x!$. We won't work through it in class; see PSE if you'd like to see the details.

In R, `dpois/ppois/qpois/rpois`.

Uniform Distribution

A **Uniform** random variable is the natural continuous analogue of "all outcomes are equally likely" — X is equally likely to fall anywhere in some interval $[a, b]$. We write $X \sim \text{Uniform}[a, b]$. We already used this distribution in a couple of earlier examples (the practice question and the transformation example in the previous set of notes both used $X \sim \text{Uniform}[0, 1]$). It's also a useful benchmark when you want to represent "no information" about where in a range a continuous outcome falls.

X is continuous with support $[a, b]$, and its pdf is

$$f_X(x) = \frac{1}{b-a}, \qquad a \leq x \leq b$$

(and $f_X(x) = 0$ outside of $[a, b]$).

```
unif_df <- data.frame(a = c(0, 0, 1), b = c(1, 2, 3))
unif_df$height <- 1 / (unif_df$b - unif_df$a)
unif_df$interval <- factor(paste0("[", unif_df$a, ",", unif_df$b, "]"))

ggplot(data = unif_df) +
  geom_segment(aes(x = a, xend = b, y = height, yend = height, color = interval),
               linewidth = 1) +
  xlab("x") +
  ylab("pdf") +
  labs(color = NULL) +
  theme_bw()
```

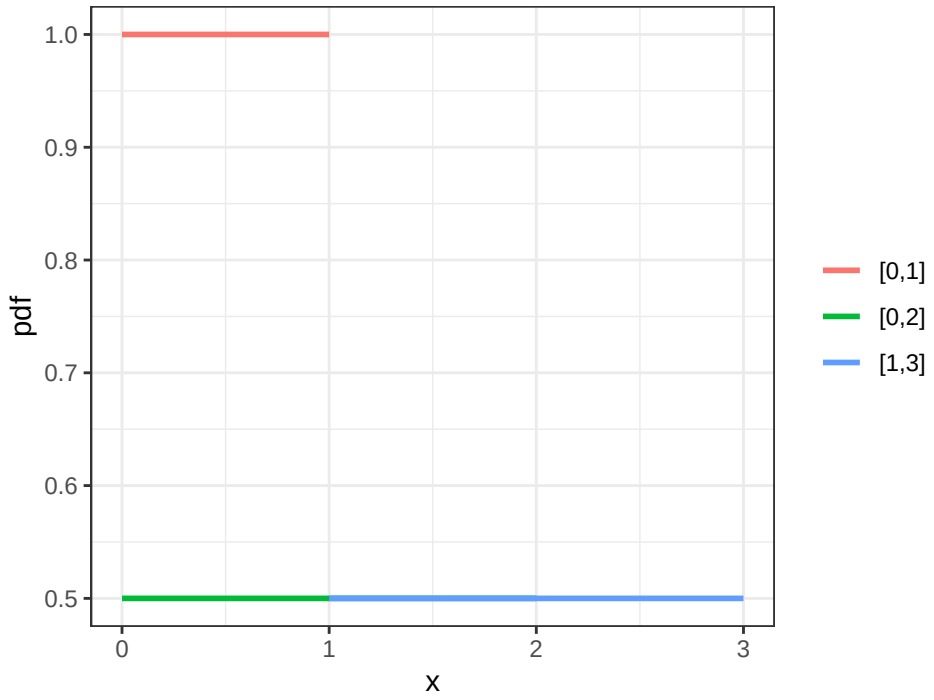

Figure 3: pdf of Uniform[a,b] for a few choices of a and b

By symmetry, $\mathbb{E}[X]$ is just the midpoint of the interval, $\mathbb{E}[X] = (a+b)/2$ (you can check this by direct integration). Similarly, $\text{var}(X) = (b-a)^2/12$.

In R, `dunif/punif/qunif/runif`.

Exponential Distribution

An **Exponential** random variable is often used to model waiting times or durations — for example, the time until a machine breaks down, the time between customer arrivals, or the length of an unemployment spell. We write $X \sim \text{Exp}(\lambda)$.

X is continuous with support $[0, \infty)$, and its pdf is

$$f_X(x) = \frac{1}{\lambda} \exp(-x/\lambda), \qquad x \geq 0,\ \lambda > 0$$

```
ggplot(data.frame(x = c(0, 10)), aes(x = x)) +
  stat_function(fun = function(x) dexp(x, rate = 1), aes(color = "lambda = 1")) +
  stat_function(fun = function(x) dexp(x, rate = 1/2), aes(color = "lambda = 2")) +
  stat_function(fun = function(x) dexp(x, rate = 1/4), aes(color = "lambda = 4")) +
  xlab("x") +
  ylab("pdf") +
  labs(color = NULL) +
  theme_bw()
```

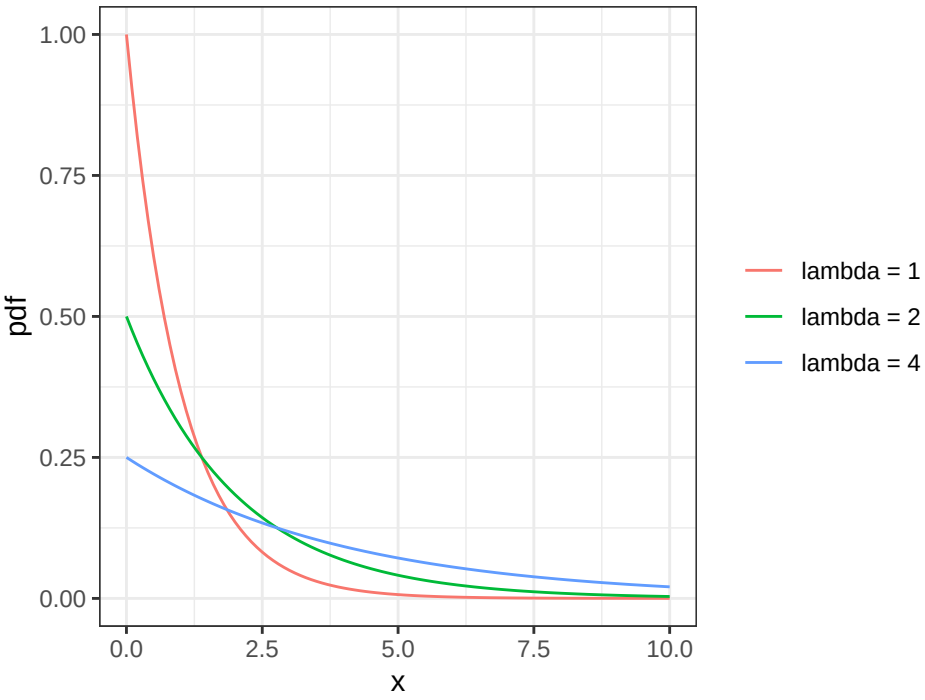

Figure 4: pdf of Exponential(lambda) for a few values of lambda

We already derived $\mathbb{E}[X] = \lambda$ for this distribution (using integration by parts) as a worked example in the previous set of notes. It also turns out that $\text{var}(X) = \lambda^2$.

In R, `dexp/pexp/qexp/rexp` — note that R parameterizes the exponential distribution by its `rate` $= 1/\lambda$ rather than by λ itself, so e.g. `dexp(x, rate = 1/lambda)` matches the pdf above.

Normal Distribution

PSE 3.12

Arguably the most important parametric distribution is the normal distribution. We will soon see that random variables following a standard normal distribution show up naturally in statistics due to the central limit theorem.

A random variable X that follows a normal distribution is continuous and the normal distribution is indexed by two parameters μ (its mean) and σ^2 (its variance). It is common to write $X \sim N(\mu, \sigma^2)$ for a random variable that follow a normal distribution. The pdf of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

```
ggplot(data.frame(x = c(-6, 10)), aes(x = x)) +
  stat_function(fun = dnorm, args = list(mean = 0, sd = 1), aes(color = "N(0,1)")) +
  stat_function(fun = dnorm, args = list(mean = 0, sd = 2), aes(color = "N(0,4)")) +
  stat_function(fun = dnorm, args = list(mean = 3, sd = 1), aes(color = "N(3,1)")) +
  xlab("x") +
```

```
  ylab("pdf") +
  labs(color = NULL) +
  theme_bw()
```

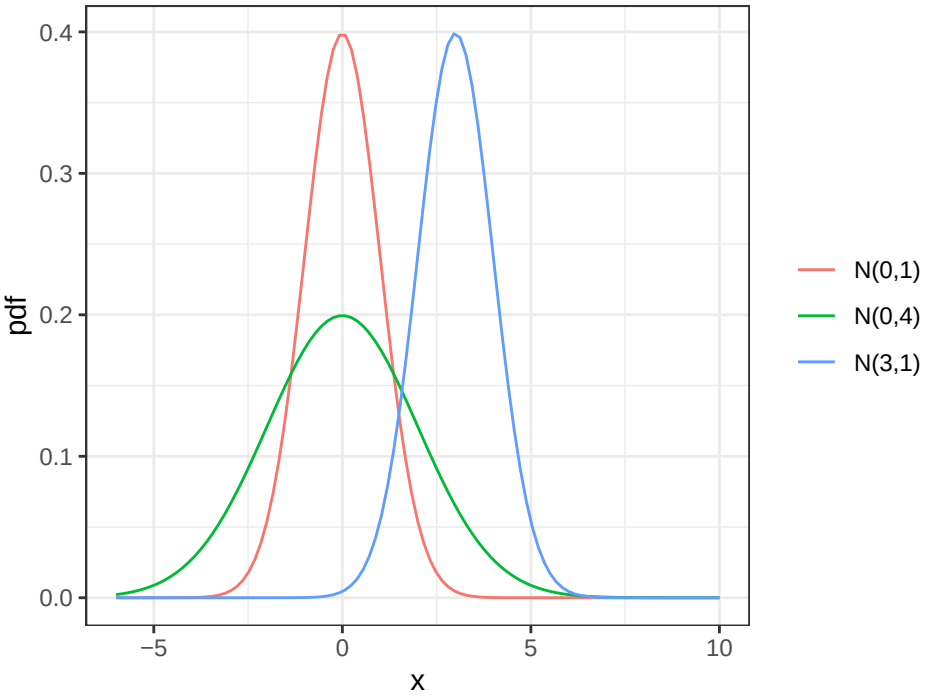

Figure 5: pdf of Normal(mu, sigma^2) for a few parameter values

A special case (but important case) is when $\mu = 0$ and $\sigma^2 = 1$. In this case, X is said to follow a **standard normal distribution**.

Another useful property of normal distributions is stated in the following propsition

Proposition: Suppose that $X \sim N(\mu, \sigma^2)$ and we define $Y = a + bX$ for some $a, b \in \mathbb{R}$. Then Y is also normally distributed as $N(a + b\mu, b^2\sigma^2)$.

A useful consequence of the previous proposition is that, when $X \sim N(\mu, \sigma^2)$, it can be "standardized" by considering the transformed random variable $Z = (X - \mu)/\sigma$. This sort of transformation will be useful to us in the statistics portion of the course.

Side-Comment

In practice we usually don't know σ and have to estimate it, so we can't actually compute $Z = (X - \mu)/\sigma$. Replacing σ with an estimate gives a quantity that is no longer exactly standard normal — it instead follows a t**-distribution**, which looks like a standard normal but with heavier tails (indexed by a "degrees of freedom" parameter reflecting how precisely σ was estimated). We'll define this formally, along with the closely related Chi-square distribution

below, in the statistics half of the course.

There is also some specialized notation that is worth mentioning for standard normal random variables. The cdf of a standard normal distribution is often denoted by Φ, and the pdf is often denoted by ϕ.

Next, if $X \sim N(\mu, \sigma^2)$, then its moment generating function is given by $M_X(t) = \exp(\mu t + \sigma^2 t^2/2)$. For a standard normal random variable, $Z \sim N(0,1)$, its moment generating function is given by $M_Z(t) = \exp(t^2/2)$.

Normal distributions arise frequently in the context of statistical inference. From your undergraduate statistics class, you may be familiar with the idea of comparing some test statistic to a critical value that comes from a normal distribution. The next table comes from Table 5.1 in the textbook and provides the a summary of the cdf of a standard normal distribution (you'll notice that some of the values of "x" in the table correspond to critical values that you may be familiar with).

Table 5.1: Normal Probabilities and Quantiles

	$\mathbb{P}[Z \leq x]$	$\mathbb{P}[Z > x]$	$\mathbb{P}[\lvert Z\rvert > x]$
$x = 0.00$	0.50	0.50	1.00
$x = 1.00$	0.84	0.16	0.32
$x = 1.65$	0.950	0.050	0.100
$x = 1.96$	0.975	0.025	0.050
$x = 2.00$	0.977	0.023	0.046
$x = 2.33$	0.990	0.010	0.020
$x = 2.58$	0.995	0.005	0.010

In R, to get values of cdfs of standard normal random variable, you can use the `pnorm` function. In R, pre-pending "p" to the name of a distribution recovers values of the cdf for a large number of distributions (e.g., `pchisq` recovers cdfs for chi-square random variables). Similarly, `qnorm` recovers quantiles for a normally distributed random variable (quantiles are the inverse of the cdf; e.g., `qnorm(.7)` would provide the 70th percentile of a standard normal random variable); `rnorm` can be used to make draws from a normally distributed random variable.

Chi-square Distribution

The **Chi-square** distribution arises directly from the normal distribution: if $Z_1, \ldots, Z_k$ are iid standard normal random variables, then

$$X = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

follows a **Chi-square distribution with k degrees of freedom**, written $X \sim \chi^2_k$. This distribution shows up throughout the statistics half of the course — for example, it's the distribution of (a scaled version of) the sample variance when sampling from a normal population, and it's the building block for the t- and F-distributions (see the side-comment below).

X is continuous with support $[0, \infty)$. Its pdf involves the Gamma function and is more complicated than we need to get into:

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \qquad x > 0$$

(deriving this requires the multivariate change-of-variables machinery for sums of random variables, which we're not covering this semester — see PSE if you're curious).

```
ggplot(data.frame(x = c(0.01, 20)), aes(x = x)) +
  stat_function(fun = dchisq, args = list(df = 1), aes(color = "k = 1")) +
  stat_function(fun = dchisq, args = list(df = 3), aes(color = "k = 3")) +
  stat_function(fun = dchisq, args = list(df = 5), aes(color = "k = 5")) +
  stat_function(fun = dchisq, args = list(df = 10), aes(color = "k = 10")) +
  xlab("x") +
  ylab("pdf") +
  labs(color = NULL) +
  theme_bw()
```

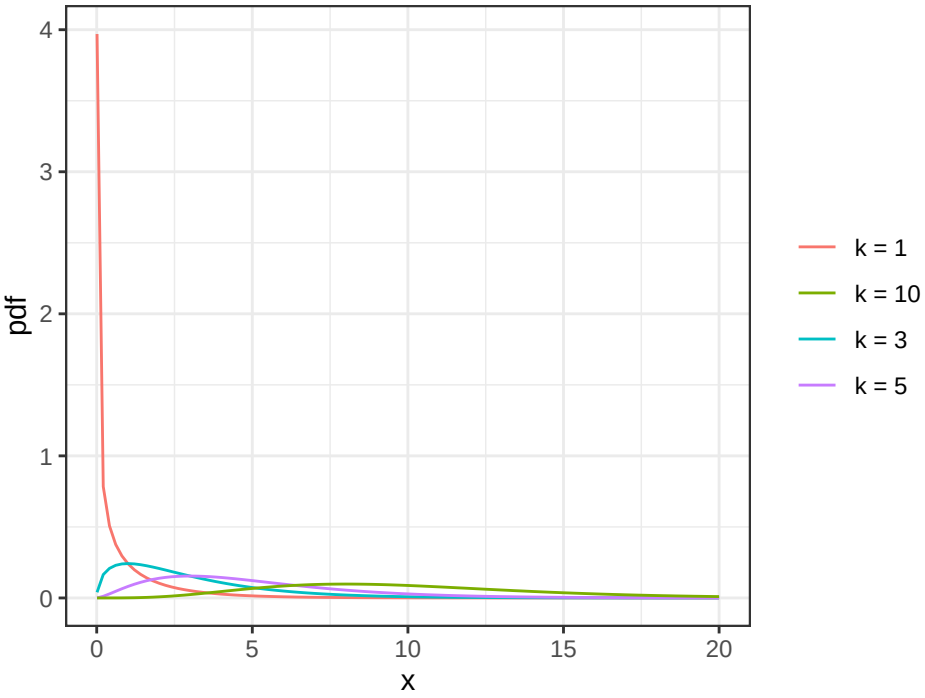

Figure 6: pdf of Chi-square(k) for a few degrees of freedom

Unlike the pdf, $\mathbb{E}[X]$ is easy to get directly from the construction above, reusing a fact we already

know: since $Z_i \sim N(0,1)$, $\mathbb{E}[Z_i^2] = \text{var}(Z_i) = 1$. So, by linearity,

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{i=1}^{k} Z_i^2\Big] = \sum_{i=1}^{k} \mathbb{E}[Z_i^2] = k$$

It also turns out that $\text{var}(X) = 2k$, though that derivation needs the fourth moment of a normal random variable, which we haven't covered — we'll just state it here.

Side-Comment

The ***F*-distribution** is the ratio of two independent Chi-square random variables, each divided by its own degrees of freedom: if $U \sim \chi^2_{k_1}$ and $V \sim \chi^2_{k_2}$ are independent, then $\frac{U/k_1}{V/k_2} \sim F_{k_1,k_2}$. It comes up when comparing two variances, or when testing multiple restrictions jointly (e.g., an F-test in a regression). We'll return to it, along with the t-distribution, in the statistics half of the course.

In R, `dchisq`/`pchisq`/`qchisq`/`rchisq`.