Monday, January 28: Distribution families

The **normal** distribution family $N(\mu, \sigma)$ is a 2-parameter family of continuous distributions, where the **mean** $\mu$ is a real number and the **standard deviation** $\sigma$ is a positive real number.

A continuous random variable $X$ is **normally distributed** with mean $\mu$ and standard deviation $\sigma$ (written $X \sim N(\mu, \sigma)$) if its probability density function is

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2}.$$
Monday, January 28: Distribution families

In R, the suffix for a normal distribution is `norm`, and its parameters are `mean` and `sd`.

The **68-95-99.7 rule** asserts that the area under a normal probability density function within 1, 2, and 3 standard deviations of the mean is approximately 0.66, 0.95, and 0.997.
The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the standard normal distribution.

The following propositions are exercises for the reader:

**Proposition** If $X$ is a random variable with $X \sim N(\mu, \sigma)$, then

$$E(X) = \mu \text{ and Var}(X) = \sigma^2.$$ 

**Proposition** If $X$ is a random variable with $X \sim N(\mu, \sigma)$, then

$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$
Normal distributions are extremely important in statistics, largely because of the **central limit theorem**:

**Proposition** (Central limit theorem) Let $X_1, X_2, \ldots$ be (reasonably nice) independent and identically distributed random variables with mean $\mu$ and standard deviation $\sigma$, and let $M_n$ be the sample mean of the first $n$ of them. Also, let $\Phi(z)$ be the cumulative distribution function of a standard normal distribution:

$$
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx
$$

Then

$$
\lim_{n \to \infty} P\left( a \leq \frac{M_n - \mu}{\sigma / \sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a)
$$

for all $a, b \in \mathbb{R}$. 
Tuesday, January 29: Distribution families

The $t$ distribution family $t(k)$ is a 1-parameter family of continuous distributions, where the **degrees of freedom** $k$ is a positive real number.

A continuous random variable $X$ is **$t$ distributed** with $k$ degrees of freedom (written $X \sim t(k)$) if its probability density function is

$$f(x) = \frac{\Gamma(((k + 1)/2))}{\Gamma(k/2)\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}.$$

The $t$ in $t$ distribution is always written in lowercase (not uppercase!)

In R, the suffix for a $t$ distribution is $t$, and its parameter is df.
Tuesday, January 29: Distribution families

The **F** distribution family $F(k, \ell)$ is a 2-parameter family of continuous distributions, where the **numerator degrees of freedom** $k$ and the **denominator degrees of freedom** $\ell$ are both positive real numbers.

A continuous random variable $X$ is **$F$ distributed** with $k$ and $\ell$ degrees of freedom (written $X \sim F(k, \ell)$) if its probability density function is

$$f(x) = \begin{cases} \frac{\Gamma((k+\ell)/2)}{\Gamma(k/2)\Gamma(\ell/2)} \left( \frac{k}{\ell} \right)^{k/2} \frac{x^{(k-2)/2}}{(1+(k/\ell)x)^{(k+\ell)/2}} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The $F$ in $F$ **distribution** is always written in uppercase (not lowercase!), except sometimes in computer software.

In R, the suffix for an $F$ distribution is `f`, and its parameters are `df1` and `df2`.
Proposition Let $Z \sim N(0, 1)$. Then $Z^2 \sim \chi^2(1)$.

Proof We know that $P(Z^2 < 0) = 0$, so let $z > 0$. Denoting the cumulative distribution function of $Z^2$ by $F$, we have

$$F(z) = P(Z^2 \leq z)$$
$$= P(-\sqrt{z} \leq Z \leq \sqrt{z})$$
$$= \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$
$$= 2 \int_{0}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt.$$ 

Let $u = t^2$. The limits of integration become 0 and $z$, and we find that

$$F(z) = \int_{0}^{z} \frac{1}{\sqrt{2\pi u}} e^{-u/2} \, du.$$
Differentiating this with respect to $z$, we get that the probability density function of $Z^2$ equals

$$f(z) = \begin{cases} \frac{1}{\sqrt{2\pi z}} e^{-z/2} & \text{if } z \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

which is the probability density function of a $\chi^2(1)$ distribution.
Corollary Let \( X \sim N(\mu, \sigma) \), and let

\[
V = \left( \frac{(X - \mu)}{\sigma} \right)^2.
\]

Then \( V \sim \chi^2(1) \).

Proof This is immediate from the previous proposition and the result that \( (X - \mu)/\sigma \sim N(0, 1) \).
Wednesday, January 30: Distribution families

Two main methods for getting at probability distributions:

1. moment generating functions
2. change of variables

Let’s look at the first of these...
Definition Let $X$ be a random variable. The moment generating function $M_X(t)$ is defined as $E(e^{tX})$.

The moment generating function need not exist, since not all moments need to exist.

Even if it exists, the moment generating function may not be finite for all $t$.

The term comes from the following proposition, which we leave as an exercise for the reader:

**proposition** Let $X$ be a random variable with moment generating function $M_X(t)$. Then

$$M_X^{(k)}(0) = E(X^k).$$
Let $\mathcal{F}$ denote the space of distributions for which all moments exist. We state the following two propositions without proof:

**Proposition** For distributions in $\mathcal{F}$, there is a one-to-one correspondence between moment generating functions and distributions.

**Proposition** A sequence of moment generating functions for distributions in $\mathcal{F}$ converges to a limiting moment generating function for a distribution in $\mathcal{F}$ if and only if its associated cumulative distributions converge to the limiting moment generating function’s associated cumulative distribution function.

We have used the second proposition to prove the central limit theorem for random variables whose distributions are in $\mathcal{F}$. 
Wednesday, January 30: Distribution families

Let $X, X_1, X_2, \ldots, X_n$ be a random variables, and let $a, b \in R$

1. If $Y = aX + b$ then

$$M_Y(t) = e^{bt} M_X(at).$$

2. If $X_1, \ldots, X_n$ are independent and $Y = X_1 + \cdots + X_n$, then

$$M_Y(t) = M_{X_1}(t) M_{X_1}(t) \cdots M_{X_n}(t).$$

We proved these in Probability Theory class.
Proposition Let $U_1, \ldots, U_k$ be independent and have $\chi^2$ distributions with $n_1, \ldots, n_k$ degrees of freedom. Then

$$U = U_1 + \cdots + U_k \sim \chi^2(n_1 + \cdots + n_k).$$

Proof Because the moment generating functions exist for $U_1, \ldots, U_k$, the moment generating function exists for their sum. Since they are independent, then

$$M_U(t) = M_{U_1}(t)M_{U_2}(t)\cdots \cdots M_{U_k}(t)$$

$$= (1 - 2t)^{-(n_1+\cdots+n_k)},$$

which is the moment generating function for a $\chi^2$ distribution with $n_1 + \cdots \cdots n_k$ degrees of freedom.

Note: We looked up the moment generating functions for $\chi^2$ distributions. It is a good exercise to compute them.
Corollary Let $X_1, \ldots, X_k$ be independent and identically distributed with distribution $N(\mu, \sigma)$. Then

$$\sum_{i=1}^{k} \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(k).$$

We leave the proof as an exercise for the reader.
Two main methods for getting at probability distributions:

1. moment generating functions
2. change of variables

Let’s look at the second of these...
The single-variable change of variables formula is given by:

**Theorem** Let \( X \) be a random variable with probability density function \( f : \mathbb{R} \rightarrow \mathbb{R} \), and suppose that \( y \) is an invertible function from a subset of \( \mathbb{R} \) onto a subset of \( \mathbb{R} \) with inverse function \( x = x(y) \). Also, let \( g : \mathbb{R} \rightarrow \mathbb{R} \) denote the probability density function of \( y(X) \). Then

\[
g(y) = f(x(y))|x'(y)|
\]

for all \( y \) in the domain of the inverse function.

For example, let \( X \sim \text{Expon}(\theta) \), so

\[
f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}
\]

What is the probability density function of \( X^2 \)?
If we let $y = x^2$, we find that for $x \geq 0$,

$$x = \sqrt{y}.$$ 

Exponential probability density functions are zero when $x < 0$, and the region $x < 0$ corresponds to the region where $y < 0$, so the previous theorem applies and gives us the probability density function of $X^2$ to be

$$g(y) = \begin{cases} 
    f(x(y))|x'(y)| & \text{for } y > 0 \\
    0 & \text{otherwise.}
\end{cases}$$

This gives us that

$$g(y) = \begin{cases} 
    \frac{1}{\theta} \frac{e^{-\sqrt{y}/\theta}}{2\sqrt{y}} & \text{for } y > 0 \\
    0 & \text{otherwise.}
\end{cases}$$
The multivariate change of variable formula is given by:

**Theorem** Let $X_1, \ldots, X_k$ be a random variables with joint probability density function $f : \mathbb{R}^k \to \mathbb{R}$. Let $y : \mathbb{R}^k \to \mathbb{R}^k$ be a function with

$$y = (y_1(x_1, \ldots, x_k), \ldots, y_k(x_1, \ldots, x_k)),$$

and suppose that $y$ is an invertible function from a subset of $\mathbb{R}^k$ onto a subset of $\mathbb{R}^k$ with inverse function $x$, so

$$x = (x_1(y_1, \ldots, y_k), \ldots, x_k(y_1, \ldots, y_k)).$$

Also, let $g : \mathbb{R}^k \to \mathbb{R}$ denote the joint probability density function of $y_1(X_1, \ldots, X_k), y_k(X_1, \ldots, X_k)$. (continued)
Friday, February 1: Distribution families

Then

$$g(y_1, \ldots, y_k) = f(x_1(y_1, \ldots, y_k), \ldots, x_k(y_1, \ldots, y_k))|\det(J)|$$

for all \((y_1, \ldots, y_k)\) in the domain of the inverse function, where

$$J = \begin{bmatrix}
\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_k} \\
\vdots & \ddots & \vdots \\
\frac{\partial x_k}{\partial y_1} & \cdots & \frac{\partial x_k}{\partial y_k}
\end{bmatrix}$$

is the Jacobian matrix of the inverse function (which gives \(x\) as a function of \(y\)).
Example Let $X_1, X_2$ be independent and identically distributed random variables, each with a Unif$(0, 1)$ distribution, so their joint probability density function is

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. What is the joint probability density function of $Y_1$ and $Y_2$?
Let

\[ y_1(x_1, x_2) = x_1 + x_2 \]
\[ y_2(x_1, x_2) = x_1 - x_2. \]

Solving this for \( x_1, x_2 \) gives

\[ x_1 = \frac{1}{2} (y_1 + y_2) \]
\[ x_2 = \frac{1}{2} (y_1 - y_2), \]

so

\[ J = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}, \]

which means that its determinant is \(-1/2\).
Using the previous theorem, we find that the joint probability density function of $Y_1$ and $Y_2$ is

$$g(y_1, y_2) = f(x_1(y_1, y_2), x_2(y_1, y_2)) \cdot \det(J)$$

$$= f\left(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right) \cdot |(-1/2)|.$$

The relevant quantities are nonzero on a square $S$ with corners at $(0, 0), (1, 1), (2, 0), (1, -1)$, so

$$g(y_1, y_2) = \begin{cases} 
1/2 & \text{if } (y_1, y_2) \in S \\
0 & \text{otherwise.}
\end{cases}$$