Outline

1) Random Samples (Normal distribution)

2) chi-squared distribution

3) t-distribution

4) F-distribution

Random Samples

- in application, many data sets consist of continuous measurements. Examples include
  - heights in centimeters of a cohort of 137 fourteen year-old boys
  - returns of 42 stocks on 18th April, 2002
  - degree to which the Catecholo-Methyltran gene is differentially expressed between cancerous and normal tissue, based on 16 microarray experiments
- it is common to model data such as these as a random sample from the normal distribution.

Random Samples

- here we study the properties of statistics (eg. sample mean and variance) of samples for which the normality assumption is reasonable

- Example:
  consider a simple experiment in which we observe a number x which has arisen through some random mechanism about which we have partial knowledge; e.g., if we observe whether or not a patient is cured by a drug, our record x of the result of the experiment can take only one of 2 values, say, 1 or 0, usually with unknown probabilities p and $1 - p$.

our model for the experiment is that we observe the value of a variable X, where $X \sim Bernoulli(p)$.

this simple experiment would be repeated on other patients and we would observe $x_1, x_2, \ldots, x_n$.

our model for the experiment is that we observe the values of independent Bernoulli(p) variables $X_1, X_2, \ldots, X_n$ - **a random sample** of size n from the Bernoulli(p) distribution

a **statistic** is any function of the X_is. In this example, the one of most interest is $X \equiv \sum_{i=1}^{n} X_i =$total number curred and $X \sim Bin(n, p)$.

Random Samples

- now consider the case in which x can take any value in an interval.

- suppose our knowledge of the random mechanism is that x can be taken to be the observed value of a variable X where,e.g., $X \sim N(\mu, \sigma^2)$ with μ, σ unknown. This means that, for any interval (a, b),

$$P(a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

  this simple experiment would be repeated and we would observe $x_1, x_2, \ldots, x_n$.

- our model for the experiment would be that we observe the values of independent $N(\mu, \sigma^2)$ variables $X_1, X_2, \ldots, X_n$- a random sample of size n from the $N(\mu, \sigma^2)$ distribution

Sampling from the Normal

- statistics of most interest now are $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$, the **sample mean** and $S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, the **sample variance**. $\bar{X}$ and S^2 are statistics which we will use to draw inferences about mean μ and the variance σ^2.

- suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $N(\mu, \sigma^2)$ distribution. Various functions of the X_is are of interest to us

- in particular, the sample mean, $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ and the sample variance, $S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Sampling from the Normal

- in order to calculate probabilities involving these functions we need to know their distributions(or **sampling distributions** - distributions arising through sampling, in this instance from the Normal distribution)

- also, any linear functions of X_i has a normal distribution: if $a_1, a_2, \ldots, a_n$ are any constants, then

$$\begin{aligned} \sum_{i=1}^n a_i X_i &\sim N\left(\sum_{i=1}^n a_i E(X_i), \sum_{i=1}^n a_i^2 Var(X_i)\right) \\ &= N\left(\mu \sum_{i=1}^n a_i, \sigma^2 \sum_{i=1}^n a_i^2\right) \end{aligned}$$

- in particular, (put $a_1 = a_2 = \ldots, a_n = 1$), $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ and put $a_1 = a_2 = \ldots, a_n = 1/n$)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

Sampling from the Normal

- functions like S^2 are nonlinear functions of the X_is and have distributions which are not Normal

- if X and Y are uncorrelated Normal variables, then X and Y are independent

- Suppose $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ variables, then $\bar{X}$ and S^2 are independent

Proof: Suppose $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ variables, then

$$\begin{aligned}
& Cov(\bar{X}, X_i - \bar{X}) \\
= \; & E(\bar{X}(X_i - \bar{X})) - E(\bar{X}) \cdot E(X_i - \bar{X}) \\
= \; & E(\bar{X} X_i) - E(\bar{X}^2), \text{ since } E(X_i - \bar{X}) = 0
\end{aligned}$$

now

$$\begin{aligned}
E(\bar{X} X_i) &= E[(\tfrac{1}{n} \textstyle\sum_{j=1}^n X_j) X_i] \\
&= \tfrac{1}{n} E(X_i(X_1 + X_2 + \ldots, + X_n)) \\
&= \tfrac{1}{n} E(X_i^2 + \textstyle\sum_{j \neq i}^n E(X_i X_j)) \\
&= \tfrac{1}{n}(\sigma^2 + \mu^2 + (n-1)\mu^2) = \sigma^2/n + \mu^2
\end{aligned}$$

thus

$$Cov(\bar{X}, X_i - \bar{X}) = \sigma^2/n + \mu^2 - E(\bar{X}^2) = 0$$

so if $\bar{X}$ is independent of $X_1 - \bar{X}$ and $X_2 - \bar{X}$ and ... and $X_n - \bar{X}$ then $\bar{X}$ is independent of $(X_1 - \bar{X})^2, (X_2 - \bar{X})^2, \ldots, (X_n - \bar{X})^2$ and also independent of $\sum_{i=1}^n (X_i - \bar{X})^2$.

Thus $\bar{X}$ is independent of $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Chi-squared distribution

- if X has density

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

  then X has the χ^2 (**chi-squared**) distribution with **degrees of freedom** v. If x has the above density, then write $X \sim \chi^2_v$

- in order to calculate the mean and variance of a χ^2_v random variable, first note that

$$\int_0^\infty e^{-x/2} x^{v/2-1} dx = 2^{v/2}\Gamma(v/2)$$

thus if $X \sim \chi_v^2$, then for any $r \geq 0$

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r \cdot \frac{e^{-x/2}x^{v/2-1}}{2^{v/2}\Gamma(v/2)}dx \\ &= \frac{1}{2^{v/2}\Gamma(v/2)}\int_0^\infty e^{-x/2}x^{\frac{(v+2r)}{2}-1}dx \\ &= \frac{2^{\frac{(v+2r)}{2}}\Gamma(\frac{v+2r}{2})}{2^{v/2}\Gamma(v/2)} \\ &= 2^r\Gamma(\frac{v}{2}+r)/\Gamma(v/2) \end{aligned}$$

Chi-squared distribution

- thus

$$E(X) = 2\Gamma(\frac{v}{2} + 1)/\Gamma(v/2) = v$$

since $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

$$E(X^2) = 2^2\Gamma(\frac{v}{2} + 2)/\Gamma(\frac{v}{2}) = v(v + 2)$$

- so the variance

$$Var(X) = v(v + 2) - v^2 = 2v$$

- **Lemma 1:** If $X_1, X_2, \ldots, X_n$ are independent χ^2 random variables with $X_i \sim \chi^2_{v_i}$, then $\sum_{i=1}^n X_i \sim \chi^2_v$, where $v = \sum_{i=1}^n v_i$

- **Lemma 2:** If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$.

Proof:

Let $V = Z^2$. Then V has cumulative distribution function

$$\begin{aligned} F_V(v) &= P(V \le v) = P(Z^2 \le v) \\ &= P(-\sqrt{v} < Z < \sqrt{v}) \\ &= \Phi(\sqrt{v}) - \Phi(-\sqrt{v}) \end{aligned}$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$ and $\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$.

Then $\Phi'(x) = \phi(x)$ and V has density $f_V(v) = \frac{\partial}{\partial v} F_V(v)$. Then

$$
\begin{aligned}
f_V(v) &= \frac{\partial}{\partial v} F_V(v) \\
&= \frac{\partial}{\partial v} \Phi(\sqrt{v}) - \frac{\partial}{\partial v} \Phi(-\sqrt{v}) \\
&= \phi(\sqrt{v}) \cdot \frac{1}{2} v^{-1/2} - \phi(-\sqrt{v}) \cdot (-\frac{1}{2} v^{-1/2}) \\
&= \frac{e^{-v/2} v^{-1/2}}{2^{1/2} \Gamma(1/2)} \sim \chi_1^2
\end{aligned}
$$

where $\Gamma(1/2) = \sqrt{\pi}$.

- thus if $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$ and $(\frac{X-\mu}{\sigma})^2 \sim \chi^2_1$.

- also if $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ random variables, then

$$\sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma})^2 \sim \chi^2_n$$

by Lemma 1.

Chi-squared distribution

- Let $X_1, \ldots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution, then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

- Let $X_1, \ldots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution. Then

$$E(S^2) = \sigma^2 \text{ and } Var(S^2) = \frac{2\sigma^4}{n-1}$$

Proof:

Since for $X \sim \chi^2_v$, $E(X) = v$ and $Var(x) = 2v$ and also

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

then

$$E(\frac{(n-1)S^2}{\sigma^2}) = n - 1 \rightarrow E(S^2) = \sigma^2$$

$$Var(\frac{(n-1)S^2}{\sigma^2}) = 2(n-1) \rightarrow Var(S^2) = \frac{2\sigma^4}{n-1}$$

The t distribution

- suppose $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution. Then the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ will subsequently be used to draw inferences about μ.

- Now $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, but σ is almost always unknown, so that it will become necessary to derive the distribution of $\frac{\bar{X}-\mu}{S/\sqrt{n}}$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance. The distribution turns out to be what is called the t distribution (or student's t distribution)

The t distribution

- Suppose $X \sim N(0,1)$ and $Y \sim \chi^2_v$ and X and Y are independent. Then $T = \frac{X}{\sqrt{Y/v}}$ has the t distribution with degree of freedom v, we write

$$T \sim t_v$$

- if $T \sim t_v$ then

$$f_T(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})}(1+\frac{t^2}{v})^{-\frac{(v+1)}{2}}, -\infty < t < \infty$$

Proof:

X and Y have joint density

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x) \cdot f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{e^{-y/2} y^{v/2-1}}{2^{v/2}\Gamma(v/2)}, \end{aligned}$$

$-\infty < x < \infty, 0 < y < \infty$.

Let $U = \frac{X}{\sqrt{Y/v}}$ and $W = \sqrt{Y}$, then $J = \frac{2w^2}{\sqrt{v}}$ and U and W have joint density

$$\begin{aligned} f_{U,W}(u,w) &= f_{X,Y}(x,y) \mid J \mid \\ &= \frac{1}{\sqrt{\pi v} 2^{\frac{v-1}{2}} \Gamma(v/2)} \cdot e^{-\frac{1}{2}(u^2/v+1)v^2} \cdot w^v \end{aligned}$$

$-\infty < u < \infty, 0 < w < \infty$.

Thus U has density

$$\begin{aligned} f_U(u) &= \frac{1}{\sqrt{\pi v}\, 2^{\frac{v-1}{2}} \Gamma(v/2)} \int_0^\infty e^{-\frac{1}{2}(u^2/v+1)v^2} \cdot w^v dw \\ &= \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\,\Gamma(v/2)} (1+u^2/v)^{-(v+1)/2}, \quad -\infty < u < \infty \end{aligned}$$

- If $T \sim t_v$, then note that $f_T(-u) = f_T(u)$ and so f_T is symmetric about 0, and $E(T) = 0$, provided that $v > 1$.

- If $T \sim t_v$, then as $v \to \infty$, T converges to a $N(0,1)$ random variable.

The t distribution

- the tables for t distribution is given in Table 5, Wackerly et al.

- if $T \sim t_v$, then $P(T \leq t_{v,\alpha}) = \alpha$. $t_{v,\alpha}$ is the αth quantile of the t_v distribution. $t_{v,1-\alpha}$ is the $1-\alpha$th quantile of the t_v distribution.

- **Example:** $t_{10,0.95} = 1.81$, $t_{15,0.99} = 2.60$

- note that $t_{\infty,0.975} = 1.96$, the upper 2.5% point of the $N(0,1)$ distribution.

The t distribution

- return now to $\frac{\bar{X}-\mu}{S/\sqrt{n}}$, where $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution and $\bar{X}$ and S^2 are the sample mean and sample variance

- if $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$, then

$$T = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{\frac{Q}{n-1}}},$$

where $Z \sim N(0,1)$ and $Q \sim \chi^2_{n-1}$

- also, the $N(0,1)$ and χ^2_{n-1} variables are independent since $\bar{X}$ and S^2 are independent. Thus from the definition of t distribution, we can conclude that $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

The t distribution

- suppose $X_1, X_2, \ldots, X_n$ are independent $N(\mu_1, \sigma_1^2)$ variables and $Y_1, Y_2, \ldots, Y_n$ are independent $N(\mu_2, \sigma_2^2)$ variables. Inferences about $\mu_1 - \mu_2$ will be based on $\bar{X} - \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i$, the difference of the sample means

- **Case 1:** Observations in pairs, $(m = n)$

  Let $Z_i = X_i - Y_i$, $i = 1, 2, \ldots, n$. Then whether or not observations are independent within pairs (they are usually not in applications) but otherwise assuming the samples are independent (Xs independent of the Ys), $Z_1, Z_2, \ldots, Z_n$ are independent $N(\mu_1 - \mu_2, \sigma_Z^2)$ random variables.

The t distribution

- now

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \bar{X} - \bar{Y}$$

and if

$$S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2,$$

then $\bar{Z}$ and S_Z^2 are independent,

$$\bar{Z} \sim N(\mu_1 - \mu_2, \frac{\sigma_Z^2}{n})$$

and

$$\frac{(n-1)S_Z^2}{\sigma_Z^2} \sim \chi^2_{n-1}$$

The t distribution

- thus

$$\frac{\frac{\bar{Z}-(\mu_1-\mu_2)}{\sigma_Z/\sqrt{n}}}{\sqrt{S_Z^2/\sigma_Z^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_Z/\sqrt{n}} \sim t_{n-1}$$

  since $N(0,1)$ and χ^2_{n-1} variables are independent

- thus inference about $\mu_1 - \mu_2$ can be based on $\bar{X} - \bar{Y}$ or equivalently on $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_Z/\sqrt{n}}$, since the unknown variances σ_1^2, σ_2^2 (and σ_Z^2) have been eliminated

The t distribution

- **Case 2:** the common variance case: $\sigma_1 = \sigma_2 = \sigma$ (σ unknown). then

$$X_1, X_2, \ldots, X_m \text{ are iid } N(\mu_1, \sigma^2)$$

$$Y_1, Y_2, \ldots, Y_n \text{ are iid } N(\mu_2, \sigma^2)$$

  and we assume that the samples are independent. Inferences about $\mu_1 - \mu_2$ will again be based on $\bar{X} - \bar{Y}$, but σ must been eliminated.

- now

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, 1)$$

The t distribution

- consider

$$S^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

where $S_X^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$=sample variance of the Xs and $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$=sample variance of the Ys. then

$$\begin{aligned} \frac{(m+n-2)S^2}{\sigma^2} &= \frac{(m-1)S_X^2}{\sigma^2} + \frac{(n-1)S_Y^2}{\sigma^2} \\ &= \chi^2_{m-1} + \chi^2_{n-1} \sim \chi^2_{m+n-2} \end{aligned}$$

since the sample variances are independent variables.

- note that $E\{\frac{(m+n-2)S^2}{\sigma^2}\} = m + n - 2$, so $E(S^2) = \sigma^2$

The t distribution

- so

$$\frac{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\sqrt{S^2/\sigma^2}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{m+n-2}}{m+n-2}}}$$

since $N(0,1)$ and χ^2_{m+n-2} are independent

- this leads to the result:

  For independent samples

$$X_1, X_2, \ldots, X_m \text{ are iid } N(\mu_1, \sigma^2)$$

$$Y_1, Y_2, \ldots, Y_n \text{ are iid } N(\mu_2, \sigma^2)$$

  we have

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

  where

$$S^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

and

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

=sample variance of the Xs

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

=sample variance of the Ys

The t distribution

- thus inferences about $\mu_1 - \mu_2$ can be based on $\bar{X} - \bar{Y}$ or equivalently on $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{m}+\frac{1}{n}}}$ since the unknown σ has now been eliminated

- note that the numerator and denominator above are independent since $\bar{X}$ is independent of S_X^2 and also independent of S_Y^2 and hence $\bar{X}$ is independent of S^2. Similarly $\bar{Y}$ is independent of S^2 and thus $\bar{X} - \bar{Y}$ is independent of S^2

The F distribution

- suppose $X_1, X_2, \ldots, X_n$ are iid $N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \ldots, Y_n$ are iid $N(\mu_Y, \sigma_Y^2)$ and the samples are independent. When comparing the variances or drawing inferences about σ_X^2/σ_Y^2 we use S_X^2/S_Y^2 (the ratio of the sample variances) and this leads us to the F distribution

- **Definition:** suppose $X \sim \chi^2_{v_1}$ and $Y \sim \chi^2_{v_2}$ and X and Y are independent. Then $F = \frac{X/v_1}{Y/v_2}$ has the F distribution with degrees of freedom v_1 and v_2. We write $F \sim F_{v_1,v_2}$.

The F distribution

- if $F \sim F_{v_1,v_2}$ then F has density function

$$f_F(u) = \frac{(\frac{v_1}{v_2})^{\frac{v_1}{2}} u^{\frac{v_1}{2}-1}(1+\frac{v_1 u}{v_2})^{-\frac{(v_1+v_2)}{2}}}{B(\frac{v_1}{2},\frac{v_2}{2})}, u > 0$$

where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

Proof:
X and Y above have joint density $f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$

$$= \frac{e^{-x/2}x^{\nu_1/2-1}}{2^{\nu_1/2}\Gamma(\nu_1/2)} \cdot \frac{e^{-y/2}y^{\nu_2/2-1}}{2^{\nu_2/2}\Gamma(\nu_2/2)}$$

Let $U = \frac{X/\nu_1}{Y/\nu_2}$ and $V = Y$. Then $J = \frac{\nu_1 v}{\nu_2}$ and U and V have joint density $f_{U,V}(u,v) = f_{X,Y}(x,y)|J|$

$$= \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \quad \frac{u^{\frac{\nu_1}{2}-1} v^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{1}{2}\left(\frac{\nu_1 u}{v_2}+1\right)v}}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \quad , u > 0, v > 0.$$

Integrating $f_{U,V}(u,v)$ over $0 < v < \infty$ we find that U has density

$$f_U(u) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} u^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1 u}{\nu_2}\right)^{-\frac{(\nu_1+\nu_2)}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \quad , u > 0$$

where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

- Tables for the F distribution can be found in Wackerly et al, table 7. Note that tables only give quantiles close to 1, this is because

$$F_{v_1,v_2,\alpha} = \frac{1}{F_{v_2,v_1,1-\alpha}}$$

Proof:
if $U = \dfrac{X/\nu_1}{Y/\nu_2}$ where $X \sim \chi^2_{\nu_1}$ and $Y \sim \chi^2_{\nu_2}$ and X and Y are independent, then $U \sim F_{\nu_1,\nu_2}$. Thus

$$\frac{1}{U} = \frac{Y/\nu_2}{X/\nu_1} \sim F_{\nu_2,\nu_1} \text{ and so}$$

$$\alpha = P(U < F_{\nu_1,\nu_2,\alpha}) = P\left(\frac{1}{U} > \frac{1}{F_{\nu_1,\nu_2,\alpha}}\right)$$

$$P\left(F_{\nu_2,\nu_1} > \frac{1}{F_{\nu_1,\nu_2,\alpha}}\right)$$

Thus $$\frac{1}{F_{\nu_1,\nu_2,\alpha}} = F_{\nu_2,\nu_1,1-\alpha}$$

or $$F_{\nu_1,\nu_2,\alpha} = \frac{1}{F_{\nu_2,\nu_1,1-\alpha}}$$

The F distribution

- **Example:**

$$F_{5,10,0.05} = \frac{1}{F_{10,5,0.95}} = 1/4.74 = 0.21$$

$$F_{5,10,0.01} = \frac{1}{F_{10,5,0.99}} = 1/10.1 = 0.10$$

- note if $X \sim N(0, 1)$, $Y \sim \chi^2_v$ and X and Y are independent, then

$$T = \frac{X}{\sqrt{Y/v}} \sim t_v$$

now $T^2 = \frac{X^2}{Y/v} = \frac{X^2/1}{Y/v}$, so $T^2 \sim F_{1,v}$ since X^2 and Y are independent.

The F distribution

- suppose now that $X_1, X_2, \ldots, X_m$ are iid $N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \ldots, Y_n$ are iid $N(\mu_Y, \sigma_Y^2)$ and the Xs and Ys are independent, i.e., the samples are independent.

- if all of the parameters are unknown, when comparing the variances or drawing inferences about σ_X^2/σ_Y^2 (for example, in trying to decide whether or not $\sigma_X = \sigma_Y$ or equivalently, $\sigma_X^2/\sigma_Y^2 = 1$, we use $\frac{S_X^2}{S_Y^2}$

The F distribution

- now

$$\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2_{m-1} \text{ and } \frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{n-1}$$

and

$$\begin{aligned} \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} &= \frac{\chi^2_{m-1}/(m-1)}{\chi^2_{n-1}/(n-1)} \\ &\quad \text{the } \chi^2 \text{ variables are independent} \\ &= \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F_{m-1,n-1} \end{aligned}$$