SAMPLING DISTRIBUTIONS AND THE CENTRAL LIMIT THEOREM

t and F distributions, plus the CLT

The t distribution

- We know that if $Y_1, Y_2, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ then
$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
- But in "real life", σ is not typically known.
- It makes sense to replace σ with an estimator s.
- As long as s is "close" to σ then the distribution of $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$ will be "close" to $N(0,1)$.

The t distribution

- Definition 5: If random variables $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$ are independent, then $T = \frac{Z}{\sqrt{W/\nu}}$ has a t DISTRIBUTION WITH ν DEGREES OF FREEDOM.
- If $Y_1, Y_2, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ then we know that $\bar{Y} \sim N(\mu, \sigma^2/n)$, that $\boxed{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}}$, and that $\bar{Y}$ is independent of s^2.
- So what is the distribution of $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$?

$$\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\cancel{(n-1)}s^2}{\sigma^2}/\cancel{(n-1)}}} \sim t_{n-1}$$

$$= \frac{\frac{\bar{Y} - \mu}{\cancel{\sigma}/\sqrt{n}}}{\frac{s}{\cancel{\sigma}}} = \frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

t distribution application

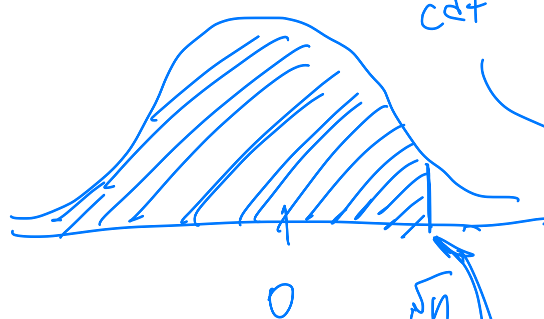
sample std. dev.

- Example 4: If observations $Y_1, Y_2, \dots, Y_n$ are iid $N(\mu, \sigma^2)$ random variables, what is the probability that $\bar{Y}$ will be no more than s units greater than the (true) mean?

μ

$$P(\bar{Y} \le \mu + s) = P(\bar{Y} - \mu \le s) = P\left(\frac{\bar{Y}-\mu}{s} \le 1\right)$$

$$= P\left(\underbrace{\frac{\bar{Y}-\mu}{s/\sqrt{n}}}_{t_{n-1}} \le \sqrt{n}\right)$$



cdf

$n = 5$

t dist

pt(sqrt(5), df = 4)

0.9555

$n = 10$?

pt(sqrt(10), df = 9)

0.9942

The F distribution

- It may be of interest to compare variances from two different normal populations, say $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.
- Given data from each population, we could calculate s_1^2 and s_2^2 and compare them.
- Definition 6: If $W_1 \sim \chi^2_{\nu_1}$ and $W_2 \sim \chi^2_{\nu_2}$ are independent, then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

has an F DISTRIBUTION WITH ν_1 NUMERATOR AND ν_2 DENOMINATOR DEGREES OF FREEDOM.

$X_1, \ldots, X_{\nu_1}$ iid $\quad s_1^2 = \frac{1}{\nu_1 - 1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$Y_1, \ldots, Y_{\nu_2}$ iid $\quad S_2^2$

indep

$$\frac{(\nu_1 - 1) s_1^2}{\sigma_1^2} \sim \chi^2_{\nu_1 - 1}$$

$$\frac{(\nu_2 - 1) s_2^2}{\sigma_2^2} \sim \chi^2_{\nu_2 - 1}$$

$$\frac{\dfrac{\cancel{(\nu_1 - 1)} s_1^2}{\sigma_1^2} \Big/ \cancel{(\nu_1 - 1)}}{\dfrac{\cancel{(\nu_2 - 1)} s_2^2}{\sigma_2^2} \Big/ \cancel{(\nu_2 - 1)}} = \frac{\dfrac{s_1^2}{\sigma_1^2}}{\dfrac{s_2^2}{\sigma_2^2}} \sim F_{\nu_1 - 1,\ \nu_2 - 1}$$

If $\sigma_1^2 = \sigma_2^2 \ (= \sigma^2)$

$$\frac{\dfrac{s_1^2}{\cancel{\sigma^2}}}{\dfrac{s_2^2}{\cancel{\sigma^2}}} \qquad \frac{s_1^2}{s_2^2} \sim F_{\nu_1 - 1,\ \nu_2 - 1}$$

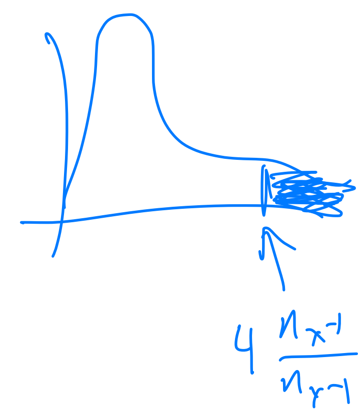
num denom

F distribution application

- Example 5: If $X_1, X_2, \dots, X_{n_X}$ are iid $N(\mu_X, \sigma^2)$ and $Y_1, Y_2, \dots, Y_{n_Y}$ are iid $N(\mu_Y, \sigma^2)$, and all the X's are independent of the Y's, what is the probability that s_Y is more than twice s_X? Calculate this probability when $n_X = 5$ and $n_Y = 8$.

$$P(s_Y > 2 s_X) = P\left(\frac{s_Y}{s_X} > 2\right) = P\left(\frac{s_Y^2}{s_X^2} > 4\right)$$

$$= P\left(\frac{s_Y^2/\sigma^2}{s_X^2/\sigma^2} > 4\right) = P\left(\frac{\frac{s_Y^2/\sigma^2}{n_Y-1}}{\frac{s_X^2/\sigma^2}{n_X-1}} > 4\left(\frac{n_X-1}{n_Y-1}\right)\right)$$



$$P\left(F > 4\left(\frac{n_X-1}{n_Y-1}\right)\right)$$

$F_{n_Y-1,\, n_X-1}$

$F \sim F_{n_Y-1,\, n_X-1}$

1 - pf(4*(5-1)/(8-1), df1 = 7, df2 = 4)

$\boxed{0.2215}$

Central Limit Theorem

don't have to be $N(\mu, \sigma^2)$

- Theorem 4: If $Y_1, Y_2, \ldots, Y_n$ are iid random variables with $E[Y_i] = \mu$ and $Var(Y_i) = \sigma^2 < \infty$, then the cdf of $U_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}}$ converges to that of a standard normal random variable, i.e.,

$Z \sim N(0,1)$

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

for all $-\infty < u < \infty$.

cdf of U_n cdf of std. normal

$$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \stackrel{\cdot}{\sim} N(0,1)$$