Sampling from the Normal Distribution

Derived distributions: Snedecor's F

Let $V_1 \sim \chi^2_{\nu_1}$ and $V_2 \sim \chi^2_{\nu_2}$ be $\boxed{\text{independent}}$ r.v.s, then the r.v.

$$X = \frac{V_1/\nu_1}{V_2/\nu_2} = \frac{\nu_2}{\nu_1}\frac{V_1}{V_2} = F_{\nu_1,\nu_2}$$

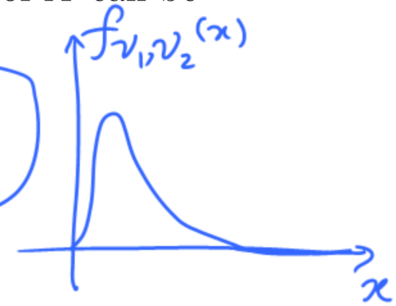
has a **Snedecor's F distribution** with ν_1 and μ_2 degrees of freedom, denoted $X \sim F_{\nu_1,\nu_2}$.

Note: $F_{\nu_1,\nu_2} \neq F_{\nu_2,\nu_1}$, $\quad F_{\nu_2,\nu_1} = \dfrac{1}{F_{\nu_1,\nu_2}}$

Snedecor named the distribution after R.A. Fisher

Using a transformation $\boxed{X = (V_1/\nu_1)/(V_2/\nu_2)}$ and $\boxed{Y = V_1/\nu_1}$, the pdf of X can be determined as

$$\longrightarrow \quad f_X(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{\frac{\nu_2}{2}-1}\left(1 + x\frac{\nu_1}{\nu_2}\right)^{-(\nu_1+\nu_2)/2}$$



$$\begin{cases} X = \dfrac{V_1/\nu_1}{V_2/\nu_2} \\ Y = \dfrac{V_1}{\nu_1} \end{cases}$$

To find $f_X(x)$:

Step1: Find $f_{X,Y}(x,y)$

Step2: Marginal $f_X(x)$ by $\int f_{X,Y}(x,y)\,dy$

$$f_{X,Y}(x,y) = f_{V_1,V_2}\left(y\nu_1, \frac{\nu_2 y}{x}\right)|J|$$

$$Y = \frac{V_1}{\nu_1} \Rightarrow \boxed{V_1 = Y\nu_1}$$

$$X = \frac{Y}{\frac{V_2}{\nu_2}} \Leftrightarrow X = \frac{\nu_2 Y}{V_2} \Leftrightarrow \boxed{V_2 = \frac{\nu_2 Y}{X}}$$

$$J = \begin{bmatrix} \dfrac{\partial V_1}{\partial x} & \dfrac{\partial V_1}{\partial y} \\ \dfrac{\partial V_2}{\partial x} & \dfrac{\partial V_2}{\partial y} \end{bmatrix}$$

Sampling from the Normal Distribution

Derived distributions: Student's t

Properties: $X = (\chi^2_{\nu_1}/\nu_1)/(\chi^2_{\nu_2}/\nu_2) \sim F_{\nu_1,\nu_2}$ $\quad \chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ are independent,

1. $\mathrm{E}X = (\nu_2/\nu_1)\mathrm{E}\chi^2_{\nu_1}/\chi^2_{\nu_2} = (\nu_2/\nu_1)\mathrm{E}(\chi^2_{\nu_1}) \cdot \mathrm{E}(1/\chi^2_{\nu_2}) \overset{?}{=} \nu_2/(\nu_2-2)$ if $\nu_2 > 2$

   $= \frac{\nu_2}{\cancel{\nu_1}}\cancel{\nu_1}\, \boxed{\mathbb{E}\left(\frac{1}{\chi^2_{\nu_2}}\right)} = \frac{\nu_2}{\nu_2-2}\,\boxed{1}$ If $\nu_2>2$

   $\mathbb{E}\left(\frac{1}{\chi^2_{\nu_2}}\right) = \frac{1}{\nu_2-2}$ Provided $\nu_2>2$

   $\mathbb{E}\left(\frac{1}{\chi^2_{\nu_2}}\right) \neq \frac{1}{\mathbb{E}(\chi^2_{\nu_2})}$

   Hint: $\mathbb{E}\left(\frac{1}{\chi^2_{\nu_2}}\right) = \int_0^\infty \frac{1}{x} f_{\chi^2_{\nu_2}}(x)\,dx$

2. if $\boxed{\nu_1 = 1}$, then

$$X = \frac{\chi^2_1/1}{\chi^2_{\nu_2}/\nu_2} \sim \left(\frac{Z}{\sqrt{\chi^2_{\nu_2}/\nu_2}}\right)^2 \sim \left(T_{\nu_2}\right)^2$$

$$F_{1,\nu_2} \overset{d}{=} \left(T_{\nu_2}\right)^2$$

3. $X^{-1} \sim (\chi^2_{\nu_2}/\nu_2)/(\chi^2_{\nu_1}/\nu_1) \sim F_{\nu_2,\nu_1}$

4. If $\nu_2 = \infty$, then $X \sim \chi^2_{\nu_1}/\nu_1$

5. $\frac{\nu_1}{\nu_2}X/(1+\frac{\nu_1}{\nu_2}X) \sim$Beta$(\nu_1/2, \nu_2/2)$, equivalent to $\chi^2_{\nu_1}/(\chi^2_{\nu_1}+\chi^2_{\nu_2}) \sim$Beta$(\nu_1/2,\nu_2/2)$

**

Motivating application: two independent normal samples

- $X_1, \ldots, X_n$ iid $N(\mu_x, \sigma^2_x)$ and $Y_1, \ldots, Y_m$ iid $N(\mu_y, \sigma^2_y)$ are independent samples
- $(n-1)S^2_x/\sigma^2_x \sim \chi^2_{n-1}$ and $(m-1)S^2_y/\sigma^2_y \sim \chi^2_{m-1}$ are then independent
- Then,

$$F = \frac{(n-1)S^2_x/\sigma^2_x}{n-1}\Big/\frac{(m-1)S^2_y/\sigma^2_y}{m-1} = \frac{\sigma^2_y}{\sigma^2_x}\frac{S^2_x}{S^2_y} \sim F_{n-1,m-1}$$

$$F_{n-1,m-1} \sim \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)} = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} = \frac{\sigma_y^2}{\sigma_x^2}\frac{S_x^2}{S_y^2}$$

- ratios of sample variances have F distribution in normal problems

STAT 542: Summary to date
Where we have been
& where we are headed

- Completed
  - Probability and random variables (definition, cdf, pdf/pmf)
  - Univariate distributions
    (definitions, expectation, transformations, families)
  - Multivariate distributions
    (joint distribution, covariance, conditional distribution, marginal distribution, independence, transformations)
  - Random samples
    (properties of sample mean, order statistics, normal sampling theory)

- Next
  - Convergence concepts (e.g., convergence in probability or in distribution)
  - Large sample results (e.g., WLLN, CLT)

Before we tackling these convergence topics, I need to mention one thing about Section 4.7 of Casella & Berger (this is out of order)

Inequalities (again)

Casella & Berger, Section 4.7

We've seen some inequalities for random variables already

1. **Markov inequality:** Suppose X is a r.v. and $g(x) \geq 0$. Then, for any $r > 0$,
$$P(g(X) \geq r) \leq \frac{\mathrm{E}g(X)}{r}.$$
2. **Jensen's inequality:** Suppose X is a r.v. and $g(x)$ is a convex function. Then,
$$\mathrm{E}g(X) \geq g(\mathrm{E}X)$$

Section 4.7 of Casella & Berger repeats Jensen's inequality & some more inequalities. The most useful additional one is the following:

Cauchy-Schwarz inequality: If X and Y are random variables with finite variances, then
$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}.$$
(Essentially, we've seen this because the correlation $|\mathrm{corr}(X,Y)| \leq 1$.)

$$|\rho_{X,Y}| \leq 1 \iff \left|\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}\right| \leq 1$$

There's also a generalization of this:

Hölder's inequality: Suppose $p, q > 0$ are positive real-numbers such that $\boxed{\frac{1}{p} + \frac{1}{q} = 1.}$ Then
$$\mathrm{E}|XY| \leq \left[\mathrm{E}|X|^p\right]^{\frac{1}{p}} \left[\mathrm{E}|Y|^q\right]^{\frac{1}{q}}$$

Convergence concepts

Introduction

$Y_n \to$ Constant as $n \to \infty$

$Y_n \to$ random variable $n \to \infty$

Motivation: $(Y_n)_{n \in \mathbb{N}}$, $n \uparrow$

- At times we can have random variables whose distributions we can't (or don't want) to work out or describe exactly.

  However, it can be possible to see that the distribution of a random variable is "close to" another distribution that *can* be described completely (often this happens for large sample sizes n).

  In which case, instead of the exact distribution of a random variable, one can use an *approximate distribution* in its place (i.e., for finding probabilities).

- Up to this point a finite-dimensional (or non-sequential) point of view
  - distribution of a single random variable X $\quad F_X(x)$
  - distribution of a random vector $(X_1, \ldots, X_n)$ $\quad F_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n)$
  - distribution of transformation $Y = g(X_1, \ldots, X_n) \longrightarrow \mathbb{P}(g(X_1, \ldots, X_n) \leq a)$

- Now consider a sequence of r.v.s Y_n, $n = 1, 2, 3$
  - distribution of Y_n may be unknown
  - distributional behavior of Y_n as $n \to \infty$ is predictable
  - example: $X_1, X_2, \ldots$ iid sequence of r.v.s with mean μ and variance σ^2
    - $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ has mean μ and variance σ^2/n (distribution generally unknown) $\quad \mathbb{E}\bar{X} = \mu,\ Var(\bar{X}) = \frac{\sigma^2}{n}$
    - "$\bar{X}_n \to \mu$" as $n \to \infty$ (law of large numbers) WLLN
    - $\bar{X}_n$ is approx $N(\mu, \sigma^2/n)$ for "large" n (central limit theorem) CLT

Convergence concepts

Introduction (cont'd)

We will consider two types of ways in which sequences of r.v.s can "converge"

$X_1, X_2, \dots$ are i.i.d

1. Convergence in distribution $X_n \xrightarrow{d} X$
2. Convergence in probability $X_n \xrightarrow{P} X$
   (a) mostly, convergence in probability to a constant (a degenerate r.v.)

$X_n \xrightarrow{a.s.} X$

In our text, there are other types of convergence ("convergence almost surely") **which we will not discuss**; these are topics discussed in STAT 6420 in more detail (the textbook attempts these but this only muddies the water).

Arguably, in statistics, convergence in distribution is most common

- conveniently, this type of convergence only requires working with probability distributions directly
- convergence in probability to a constant is the same as convergence in distribution to a constant (as we will see)

Set up: Let $Y_1, Y_2, \dots, Y_n, \dots$ be a sequence of r.v.s

- the distribution of Y_n can change with n
- One common situation
  1. $X_1, X_2, \dots$ is an iid sequence
  2. $Y_n = g(X_1, \dots, X_n)$ for each n
  3. functions $g()$ include: mean, sample variance, minimum, maximum