STAT 712 fa 2022 Lec 12 slides

Random samples, statistics, pivotal quantities

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Random sample

A collection of independent rvs with the same distribution is a *random sample*.

- Often denote by $X_1, \ldots, X_n$, where n is the *sample size*.
- In random sample, $X_1, \ldots, X_n$ are *iid: independent and identically distributed*.
- Common distribution of $X_1, \ldots, X_n$ called the *population distribution*.
- Sometimes write $X_1, \ldots, X_n \overset{\text{ind}}{\sim} f_X$, where f_X is the pop. pdf/pmf.
- Joint pdf/pmf of $X_1, \ldots, X_n \overset{\text{ind}}{\sim} f_X$ given by $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_X(x_i)$.

Goal is to learn from $X_1, \ldots, X_n$ about the population distribution.

Sample statistic

A *statistic* is any function of the rvs in the random sample.

Setup: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} f_X$, and consider the rv

$$T_n := T(X_1, \dots, X_n)$$

for some function $T : \mathbb{R}^n \to \mathbb{R}$.

Sampling distribution

The distribution of a statistic is called the *sampling distribution* of the statistic.

- We want to learn about the population distribution from sample statistics.
- What we can learn from a statistic depends on its sampling distribution.

Theorem (Moment results for mean and variance)

Let $X_1, \ldots, X_n$ be a rs from a dist. with mean μ and variance $\sigma^2 < \infty$. Then for

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

we have $\mathbb{E}\bar{X}_n = \mu$, $\operatorname{Var}\bar{X}_n = \sigma^2/n$, and $\mathbb{E}S_n^2 = \sigma^2$.

Exercise: Prove the above.

Recall:

Theorem (mgf method for sums of independent rvs)

Let $X_1, \dots, X_n$ be iid with mgf M_X. Then the mgf of $Y = \sum_{i=1}^n X_i$ is given by

$$M_Y(t) = [M_X(t)]^n.$$

Exercises: Find the sampling distribution of $\bar{X}_n$ when

1. $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$.
2. $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$.

Pivotal quantity

A *pivotal quantity* is a function of sample statistics and population parameters which has a known distribution (More rigorous treatment coming in STAT 713).

Pivotal quantities are useful for constructing confidence intervals.

Confidence interval

A $(1-\alpha)\times 100\%$ *confidence interval* for a parameter θ is an interval (L, U), where L and U are random variables such that $P(L < \theta < U) \geq 1-\alpha$.

Exercise: If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \text{Normal}(0, 1).$$

Use this result to construct a $(1-\alpha)\times 100\%$ confidence interval for μ.

Standard Normal distribution:

- The pdf of the Normal$(0,1)$ distribution is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \text{for } z \in \mathbb{R}.$$

- Denote by Φ the cdf: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ for $z \in \mathbb{R}$.
- mgf: $M_Z(t) = e^{t^2/2}$
- Let z_ξ satisfy $P(Z > z_\xi) = \xi$, where $Z \sim$ Normal$(0,1)$.

Chi-squared distributions:

- The pdf of the χ^2_ν distribution is given by

$$f_W(w;\nu) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} w^{\nu/2-1} \exp\left[-\frac{w}{2}\right], \quad \text{for } x > 0.$$

- ν is called the *degrees of freedom*
- mgf: $M_W(t) = (1-2t)^{-\nu/2}$ for $t < 1/2$.
- Let $\chi^2_{\nu,\xi}$ satisfy $P(W > \chi^2_{\nu,\xi}) = \xi$, where $W \sim \chi^2_\nu$.

Let $Z_1, \dots, Z_\nu \overset{\text{ind}}{\sim} \text{Normal}(0,1)$, then

$$W = Z_1^2 + \cdots + Z_\nu^2 \sim \chi^2_\nu.$$

Exercise: Prove the above.

t distributions:

- The pdf of the t_ν distribution is given by

$$f_T(t;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{\sqrt{\nu\pi}}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad \text{for } t \in \mathbb{R}.$$

- ν is called the *degrees of freedom*
- mgf: does not exist!
- Let $t_{\nu,\xi}$ satisfy $P(T > t_{\nu,\xi}) = \xi$, where $T \sim t_\nu$.

Let Z and W be independent rvs such that $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$, then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$

Exercise: Prove above via finding the joint density of (T, U), where $U = W$.

F distributions:

- The pdf of the F_{ν_1,ν_2} distribution is given by

$$f_R(r;\nu_1,\nu_2) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} r^{(\nu_1-2)/2}\left(1+\frac{\nu_1}{\nu_2}r\right)^{-(\nu_1+\nu_2)/2}$$

  for $r > 0$.

- ν_1 and ν_2 are called the *numerator and denominator degrees of freedom*.
- mgf: does not exist!
- Let $F_{\nu_1,\nu_2,\xi}$ satisfy $P(R > F_{\nu_1,\nu_2,\xi}) = \xi$, where $R \sim F_{\nu_1,\nu_2,\xi}$.

Let W_1 and W_2 be independent rvs such that $W_1 \sim \chi^2_{\nu_1}$ and $W_2 \sim \chi^2_{\nu_2}$, then

$$R = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2}.$$

Exercise: Prove above via finding the joint density of (R, U), where $U = W_2$.

More about the F distributions

1. If $R \sim F_{\nu_1,\nu_2}$ then $1/R \sim F_{\nu_2,\nu_1}$.
2. If $T \sim t_\nu$ then $T^2 \sim F_{1,\nu}$
3. If $R \sim F_{\nu_1.\nu_2}$ then $\dfrac{(\nu_1/\nu_2)R}{1+(\nu_1/\nu_2)R} \sim \text{Beta}(\nu_1/2, \nu_2/2)$.

Exercise: Prove the above.

Theorem (Pivotal quantity results with sample from Normal)

If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, *then*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1), \qquad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}, \qquad \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Exercise: Derive the above. For the third one, we need the following result:

Theorem (Independence of $\bar{X}_n$ and S^2 when population is Normal)

If $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$ *then* $\bar{X}_n \perp\!\!\!\perp S_n^2$.

Exercise: Prove the above.

Non-central t distributions:

- The pdf of the $t_{\nu,\phi}$ distribution is given by

$$f_T(t;\nu,\phi) = \frac{e^{-\phi^2/2}}{\sqrt{\pi\nu}\Gamma(\nu/2)} \sum_{k=0}^{\infty} \frac{(2/\nu)^{k/2}(\phi t)^k}{k!} \frac{\Gamma((\nu+k+1)/2)}{(1+(t^2/\nu))^{(\nu+k+1)/2}}, \quad t \in \mathbb{R}.$$

- ν is called the *degrees of freedom*
- ϕ is called the *non-centrality parameter*
- mgf: does not exist!

For $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$ with $Z \perp\!\!\!\perp W$, for any $\phi \in \mathbb{R}$, we have

$$T = \frac{Z+\phi}{\sqrt{W/\nu}} \sim t_{\nu,\phi}.$$

Exercise: For $X_1,\dots,X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu,\sigma^2)$, show that

$$\sqrt{n}(\bar{X}_n - \mu_0)/S_n \sim t_{n-1,\phi}, \qquad \text{with } \phi = \sqrt{n}(\mu-\mu_0)/\sigma.$$

Non-central chi-squared distributions:

- The pdf of the $\chi^2_\nu(\phi)$ distribution is given by

$$f_W(w;\nu,\phi) = \sum_{k=0}^{\infty} \frac{e^{-\phi/2}(\phi/2)^k}{k!} \frac{1}{\Gamma(\frac{\nu+2k}{2})2^{\frac{\nu+2k}{2}}} w^{\frac{\nu+2k}{2}-1} e^{-w/2}, \quad \text{for } w > 0.$$

- ν is called the *degrees of freedom*
- ϕ is called the *non-centrality parameter*
- mgf: $M_W(t) = \exp(\phi t/(1-2t))(1-2t)^{-\nu/2}$ for $t < 1/2$.

Let $Z_1, \dots, Z_\nu \overset{\text{ind}}{\sim} \text{Normal}(0,1)$, then

$$W = (Z_1 + \mu_1)^2 + \cdots + (Z_\nu + \mu_\nu)^2 \sim \chi^2_\nu(\phi = \mu_1^2 + \cdots + \mu_\nu^2).$$

Exercise: Prove the above.

Non-central F distributions:

- The pdf of the $F_{\nu_1,\nu_2}(\phi)$ distribution is given by

$$f_R(r;\nu_1,\nu_2) = \sum_{k=0}^{\infty} \frac{e^{-\phi/2}(\phi/2)^2}{k!} \frac{\Gamma(\frac{\nu_1+\nu_2+2k}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_1+2k}{2})} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1+2k}{2}} r^{\frac{\nu_1+2k}{2}-1}}{\left(1+r\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1+\nu_2+2k}{2}}}, \quad r > 0.$$

- ν_1 and ν_2 are called the *numerator and denominator degrees of freedom*.
- ϕ is the *non-centrality parameter*.
- mgf: does not exist!

Let W_1 and W_2 be independent rvs such that $W_1 \sim \chi^2_{\nu_1}(\phi)$ and $W_2 \sim \chi^2_{\nu_2}$, then

$$R = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2}(\phi).$$

Exercise: Prove above via finding the joint density of (R, U), where $U = W_2$.

1. Random samples and sample statistics
2. Normal pivotal quantity results
3. Non-central distributions
4. Miscellaneous and two-sample pivotal quantity results

Linear combinations of independent Normal rvs

Let $X_1, \dots, X_n$ be rvs such that $X_j \sim \text{Normal}(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$ and let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ be constants. Then for the rvs

$$
\begin{aligned}
U &= a_1 X_1 + \cdots + a_n X_n \\
V &= b_1 X_1 + \cdots + b_n X_n,
\end{aligned}
$$

we have $\text{Cov}(U, V) = \sum_{j=1}^n a_j b_j \sigma_j^2$ and

$$U \perp\!\!\!\perp V \iff \text{Cov}(U, V) = 0.$$

Exercise: Let $X_1, X_2 \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$ and let

$$
\begin{aligned}
U &= a_1 X_1 + a_2 X_2 \\
V &= b_1 X_1 + b_2 X_2.
\end{aligned}
$$

Show that $U \perp\!\!\!\perp V \iff \text{Cov}(U, V) = 0$.

Theorem (Pivotal quantity results with two Normal samples)

For two independent random samples

$$X_1, \dots, X_{n_1} \overset{\text{ind}}{\sim} \text{Normal}(\mu_1, \sigma_1^2), \quad \bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \quad S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$$

$$Y_1, \dots, Y_{n_2} \overset{\text{ind}}{\sim} \text{Normal}(\mu_2, \sigma_2^2), \quad \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$$

we have

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{Normal}(0, 1)$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\text{approx}}{\sim} t_{\hat{\nu}} \qquad \textit{(discuss later)}$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1 - 1, n_2 - 1}$$

Exercise: Derive the above.