Derivation of Student's T Distribution I

Part I: The Chi Square Distribution and related results

Bruce M. Boghosian

Tufts University School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

Using the sample variance for estimation

Bruce M. Boghosian

- Recall difference between *variance* and *sample variance*
  - $\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^n \left(Y_j - \overline{Y}\right)^2$
  - $S_Y^2 = \frac{1}{n-1}\sum_{j=1}^n \left(Y_j - \overline{Y}\right)^2$
  - where $\overline{Y} := \frac{1}{n}\sum_{k=1}^n Y_k$
- By the CLT, $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is normally distributed.
- Question: Is $T = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ also normally distributed?
- Answer: For very large n, there is little difference in distributions of Z and T.

Large samples and small samples

- For many years it was believed T was also normally distributed for small n.
- William Sealy Gossett (1876-1937) was the first to realize that it was not.
- Small-sample quality assurance at the Guinness brewery in Dublin
  - Barley and other ingredients came in small batches (small n) from small farms.
  - σ was generally unknown and had to be inferred from the data
  - Gossett noticed distributions of $T = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ decayed slower than a normal pdf.
  - Distribution was still bell-shaped, but the tails were "thicker".

Outline of methodology

- Review gamma distributions and sums of gamma-distributed random variables
- Understand chi square distribution as a special case of the gamma distribution
- Show sums of squares of iid normal r.v.s, $\sum_{j=1}^{n} Z_j^2$, are chi square distributed
- Show $\overline{Y}$ and S_Y^2 are independent
- Show $\frac{(n-1)S^2}{\sigma^2}$ is chi square distributed
- Derive pdf of ratio of two iid chi square r.v.s, which is called an F distribution
- Show that $T^2 = \left(\frac{\overline{Y}-\mu}{S/\sqrt{n}}\right)^2$ is F distributed
- Use the above to derive the T distribution pdf f_T

The gamma function

- Definition of the gamma function: $\boxed{\Gamma(r) := \int_0^\infty du\ e^{-u} u^{r-1}}$
- Special cases (first by substitution $u = w^2$, second is elementary):

$$\boxed{\Gamma\left(1/2\right) = \int_0^\infty du\ \frac{e^{-u}}{\sqrt{u}} = \sqrt{\pi} \quad \text{and} \quad \Gamma(1) = \int_0^\infty du\ e^{-u} = 1.}$$

- Recurrence formula

$$\Gamma(r+1) := \int_0^\infty du\ e^{-u} u^r = -e^{-u} u^r \Big|_0^\infty + r \int_0^\infty du\ e^{-u} u^{r-1}$$

$$\boxed{\Gamma(r+1) = r\Gamma(r)}$$

- Then $\Gamma(2) = 1 \cdot \Gamma(1) = 1!$, $\Gamma(3) = 2 \cdot \Gamma(2) = 2!$, and more generally

$$\boxed{\Gamma(r+1) = r! \quad \text{if } r \in \mathbb{Z}}$$

The beta function

Bruce M. Boghosian

- Definition

$$\boxed{B(r,s) := \int_0^1 dt\ t^{r-1}(1-t)^{s-1}}$$

- Symmetry

$$\boxed{B(r,s) = B(s,r)}$$

- Relationship to gamma function

$$\boxed{B(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}}$$

Proof of relationship between beta and gamma functions

Bruce M. Boghosian

- First note

$$\Gamma(r)\Gamma(s) = \int_0^\infty du\; e^{-u} u^{r-1} \int_0^\infty dv\; e^{-v} v^{s-1} = \int_0^\infty du \int_0^\infty dv\; e^{-u-v} u^{r-1} v^{s-1}$$

- Change variables $u = zt$ and $v = z(1-t)$, so $z = u + v$ and $t = u/(u+v)$
- Jacobian is $J = \begin{vmatrix} t & z \\ 1-t & -z \end{vmatrix} = z$

$$\begin{aligned}
\Gamma(r)\Gamma(s) &= \int_0^\infty dz \int_0^\infty dt\; z\, e^{-z} (zt)^{r-1} \left[z(1-t)\right]^{s-1} \\
&= \int_0^\infty dz\; e^{-z} z^{r+s-1} \int_0^1 dt\; t^{r-1} (1-t)^{s-1} \\
&= \Gamma(r+s) B(r,s),
\end{aligned}$$

from which the desired result immediately follows. $\square$

The gamma pdf I

- Definition of the two-parameter gamma pdf for $y \geq 0$

$$\boxed{f_Y(y) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}}$$

- Normalization follows immediately from definition of gamma function
- Expectation:

$$E(Y) = \int_0^\infty dy\ \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} y = \frac{r}{\lambda} \int_0^\infty dy\ \frac{\lambda^{r+1}}{\Gamma(r+1)} y^{(r+1)-1} e^{-\lambda y}$$

$$\boxed{E(Y) = \frac{r}{\lambda}}$$

The gamma pdf II

- Mean sq.:

$$E(Y^2) = \int_0^\infty dy\, \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} y^2 = \frac{r(r+1)}{\lambda^2} \int_0^\infty dy\, \frac{\lambda^{r+2}}{\Gamma(r+2)} y^{r+1} e^{-\lambda y}$$

$$\boxed{E(Y^2) = \frac{r(r+1)}{\lambda^2}}$$

- Variance:

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2$$

$$\boxed{\mathrm{Var}(Y) = \frac{r}{\lambda^2}}$$

The beta pdf I

- Definition of the two-parameter beta pdf for $y \geq 0$

$$\boxed{f_\Theta(\theta) = \frac{1}{B(r,s)}\theta^{r-1}(1-\theta)^{s-1}}$$

- Normalization follows immediately from definition of beta function
- Expectation:

$$E(\Theta) = \frac{B(r+1,s)}{B(r,s)} = \frac{\Gamma(r+1)\Gamma(s)}{\Gamma(r+s+1)} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} = \frac{r\Gamma(r)\Gamma(r+s)}{(r+s)\Gamma(r+s)\Gamma(r)}$$

$$\boxed{E(\Theta) = \frac{r}{r+s}}$$

The beta pdf II

- Mean square:

$$E(\Theta^2) = \frac{B(r+2,s)}{B(r,s)} = \frac{\Gamma(r+2)\Gamma(s)}{\Gamma(r+s+2)} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} = \frac{r(r+1)\Gamma(r)\Gamma(r+s)}{(r+s)(r+s+1)\Gamma(r+s)\Gamma(r)}$$

$$\boxed{E(\Theta^2) = \frac{r(r+1)}{(r+s)(r+s+1)}}$$

- Variance:

$$\text{Var}(\Theta) = E(\Theta^2) - [E(\Theta)]^2 = \frac{r(r+1)}{(r+s)(r+s+1)} - \left(\frac{r}{r+s}\right)^2$$

$$\boxed{\text{Var}(\Theta) = \frac{rs}{(r+s)^2(r+s+1)}}$$

Sums of random variables

- Suppose $U, V \in \mathbb{R}$ are independent r.v.s with pdfs $f_U(u)$ and $f_V(v)$
- We wise to find the pdf of the sum $U + V$. Begin with the cdf

$$\begin{aligned} F_{U+V}(t) =& P(U + V < t) \\ =& \int_{-\infty}^{+\infty} du \int_{-\infty}^{t-u} dv \; f_U(u) f_V(v). \end{aligned}$$

- Differentiating both sides with respect to t yields

$$\boxed{f_{U+V}(t) = \int_{-\infty}^{+\infty} du \; f_U(u) f_V(t - v).}$$

Sums of gamma-distributed r.v.s

- Suppose U and V are independent gamma-distributed r.v.s with parameters (r, λ) and $(s.\lambda)$, respectively.
- Then $f_{U+V}(t)$ is given by a convolution

$$f_{U+V}(t) = \int_{-\infty}^{+\infty} du\ f_U(u) f_V(t-u) = \int_0^t du\ \left[\frac{\lambda^r}{\Gamma(r)} u^{r-1} e^{-\lambda u}\right] \left[\frac{\lambda^s}{\Gamma(s)} (t-u)^{s-1} e^{-\lambda(t-u)}\right]$$

$$= e^{-\lambda t} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} \int_0^t du\ u^{r-1}(t-u)^{s-1} = e^{-\lambda t} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} t\ t^{r-1} t^{s-1} \int_0^1 dz\ z^{r-1}(1-z)^{s-1}$$

$$= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} t^{r+s-1} B(r,s) e^{-\lambda t} = \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} t^{r+s-1} \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} e^{-\lambda t}$$

- So $U + V$ is also gamma-distributed, with parameters $(r + s, \lambda)$,

$$\boxed{f_{U+V}(t) = \frac{\lambda^{r+s}}{\Gamma(r+s)} t^{r+s-1} e^{-\lambda t}}$$

The chi squared distribution I

- **Thm.:** Let $U = \sum_{j=1}^{n} Z_j^2$ where the Z_j are iid standard normal.
- Then U is gamma-distributed with parameters $r = {}^{n}/_{2}$ and $\lambda = {}^{1}/_{2}$,

$$f_U(u) = \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} u^{(n/2)-1} e^{-u/2} \qquad \text{where } u > 0$$

- **Pf.:** First take $n = 1$. For all $u \geq 0$,

$$F_{Z^2}(u) = P(Z^2 \leq u) = P\left(-\sqrt{u} \leq Z \leq +\sqrt{u}\right) = 2P(0 \leq Z \leq \sqrt{u}).$$

or

$$F_{Z^2}(u) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} dz\; e^{-z^2/2}$$

The chi squared distribution II

Bruce M. Boghosian

- **Pf. (continued):** We have

$$F_{Z^2}(u) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} dz\ e^{-z^2/2}$$

- Differentiate to find that Z is gamma-distributed with parameters $r = {}^1\!/_2$ and $\lambda = {}^1\!/_2$,

$$f_{Z^2}(u) = \frac{2}{\sqrt{2\pi}\sqrt{u}} e^{-u/2} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)} u^{(1/2)-1} e^{-u/2}$$

- Hence, if $U = \sum_{j=1}^n Z_j^2$, it must be that U is gamma-distributed with parameters $r = {}^n\!/_2$ and $\lambda = {}^1\!/_2$.

The chi squared distribution III

- **Def.:** The pdf of $U = \sum_{j=1}^{n} Z_j^2$, where Z_j are iid standard normal, is called the *chi squared distribution with n degrees of freedom*.

$$\boxed{f_{Z^2}(u) = \frac{\left(\frac{n}{2}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)} u^{(n/2)-1} e^{-u/2}}$$

- The chi squared distribution is a special case of the gamma distribution with parameters $n/2$ and $1/2$.

Orthogonal matrices I

- Column vector $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$
- Transpose is row vector $v^T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$
- Square of length of vector is given by

$$\|v\|^2 = v_1^2 + \cdots + v_n^2 = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v^T v$$

- Hence length of vector is given by

$$\|v\| = \sqrt{v^T v}$$

Orthogonal matrices II

- Consider a linear transformation of the vector, $u = Av$, where A is an $n \times n$ matrix, and demand that it preserve length

$$0 = u^T u - v^T v = (Av)^T (Av) - v^T v = v^T A^T A v - v^T v = v^T (A^T A - I) v.$$

- If we require the above to be true for all vectors v, it must be that

$$\boxed{A^T A = I}$$

- A matrix with this property is called an *orthogonal matrix*.
- If the squarematrix A is nonsingular, postmultiplying both sides of the the above by A^{-1} yields

$$\boxed{A^T = A^{-1}.}$$

and premultiplying both sides of this by A yields

$$\boxed{AA^T = I}$$

Orthogonal matrices III

- Let a_j denote the jth column of A
- The equation $A^T A = I$ indicated

$$a_j^T a_k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

- Hence the rows and columns of an orthogonal matrix are unit vectors.

Orthogonal matrices IV

Bruce M. Boghosian

- Given an orthogonal matrix A, we have $A^T A = I$
- Take the determinant of both sides and use the theorems on determinants
  - $\det(AB) = \det(A)\det(B)$
  - $\det(A^T) = \det(A)$
- The result is

$$\det(A^T A) = \det(A^T)\det(A) = [\det(A)]^2 = \det(I) = 1$$

and hence

$$\boxed{\det(A) = \pm 1.}$$

- A transformation with $\det(A) = +1$ is a *proper orthogonal transformation*
- A transformation with $\det(A) = -1$ is an *improper orthogonal transformation*

Orthogonal matrices V

- The transformation $u = Av$ can be written $u_i = \sum_{j=1}^{n} A_{ij} v_j$
- The (i,j)th element of the Jacobian matrix is A_{ij} so the matrix is A,

$$\frac{\partial u_i}{\partial v_j} = A_{ij}$$

- Jacobian factor for transforming n-dimensional integral over the v is

$$J = |\det(A)| = |\pm 1| = 1.$$

- Hence if we write $du = du_1 \cdots du_n$ and $dv = dv_1 \cdots dv_n$,

$$\int du\ f(u) = \int dv\ f(Av).$$

The transformation from X to Z I

- Let $X_j = \frac{Y_j - \mu}{\sigma}$ for $j = 1, \dots, n$
- We know the X_j are $N(0,1)$ (standard normal)
- Let A be an orthogonal matrix whose last row is $\begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{bmatrix}$
- Then

$$Z_n = \frac{X_1}{\sqrt{n}} + \cdots + \frac{X_n}{\sqrt{n}} = \sqrt{n}\,\overline{X}$$

The transformation from X to Z II

Bruce M. Boghosian

- Also, for variables z and x with $z = Ax$, we have

$$\|z\|^2 = z_1^2 + \cdots + z_n^2 = x_1^2 + \cdots + x_n^2 = \|x\|^2$$

- Since $\|x\|^2 = \|z\|^2$ and the Jacobian is one, the multivariate pdfs transform as

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) =& (2\pi)^{-n/2} \exp\left[-\frac{1}{2}(x_1^2 + \cdots + x_n^2)\right] \\ =& (2\pi)^{-n/2} \exp\left[-\frac{1}{2}(z_1^2 + \cdots + z_n^2)\right] = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) \end{aligned}$$

- Hence the Z_j are also iid $N(0,1)$ (standard normal) r.v.s

The transformation from X to Z III

- Finally we note that

$$\sum_{j=1}^{n} Z_j^2 = \sum_{j=1}^{n-1} Z_j^2 + n\overline{X}^2 = \sum_{j=1}^{n} X_j^2 = \sum_{j=1}^{n} (X_j - \overline{X})^2 + n\overline{X}^2$$

- Hence we have

$$\frac{1}{n-1}\sum_{j=1}^{n-1} Z_j^2 = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \overline{X})^2 = S^2$$

- Hence S^2 is independent of $\overline{X}^2$, and hence of $\overline{X}$.
- Since $Y_j = \mu + \sigma X_j$, S^2 is independent of $\overline{Y}$. $\square$

Summary

- We have provided some background and motivation for small-sample statistics
- We have reviewed the gamma and beta functions
- We have reviewed the gamma and beta pdfs
- We have learned about the chi squared distribution
- We have reviewed the linear algebra of orthogonal matrices
- We have proven the independence of $\overline{Y}$ and S^2