Distributions Derived From the Normal Distribution

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

1. Distributions Derived from Normal Random Variables
   - χ^2, t, and F Distributions
   - Statistics from Normal Samples

Normal Distribution

Definition. A **Normal / Gaussian** random variable $X \sim N(\mu, \sigma^2)$ has density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < +\infty.$$

with mean and variance parameters:

$$\begin{array}{rcl} \mu & = & E[X] & = & \int_{-\infty}^{+\infty} x f(x) dx \\ \sigma^2 & = & E[(X-\mu)^2] & = & \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \end{array}$$

Note: $-\infty < \mu < +\infty$, and $\sigma^2 > 0$.

Properties:

- Density function is symmetric about $x = \mu$.
  $$f(\mu + x^*) = f(\mu - x^*).$$
- $f(x)$ is a maximum at $x = \mu$.
- $f''(x) = 0$ at $x = \mu + \sigma$ and $x = \mu - \sigma$
  (inflection points of bell curve)
- Moment generating function:
  $$M_X(t) = E[e^{tX}] = e^{\mu t + \sigma^2 t^2/2}$$

Chi-Square Distributions

Definition. If $Z \sim N(0,1)$ (Standard Normal r.v.) then

$$U = Z^2 \sim \chi_1^2,$$

has a **Chi-Squared distribution with 1 degree of freedom.**

Properties:

- The density function of U is:
  $$f_U(u) = \frac{u^{-1/2}}{\sqrt{2\pi}} e^{-u/2}, \quad 0 < u < \infty$$
- Recall the density of a $Gamma(\alpha, \lambda)$ distribution:
  $$g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \ x > 0,$$
  So U is $Gamma(\alpha, \lambda)$ with $\alpha = 1/2$ and $\lambda = 1/2$.
- Moment generating function
  $$M_U(t) = E[e^{tU}] = [1 - t/\lambda]^{-\alpha} = (1 - 2t)^{-1/2}$$

Chi-Square Distributions

Definition. If $Z_1, Z_2, \ldots, Z_n$ are i.i.d. $N(0,1)$ random variables

$$V = Z_1^2 + Z_2^2 + \ldots Z_n^2$$

has a **Chi-Squared distribution with n degrees of freedom.**

Properties (continued)

- The Chi-Square r.v. V can be expressed as:
$$V = U_1 + U_2 + \cdots + U_n$$
where $U_1, \ldots, U_n$ are i.i.d χ_1^2 r.v.
- Moment generating function
$$M_V(t) = E[e^{tV}] = E[e^{t(U_1+U_2+\cdots+U_n)}]$$
$$= E[e^{tU_1}] \cdots E[e^{tU_n}] = (1-2t)^{-n/2}$$
- Because U_i are i.i.d. $Gamma(\alpha = 1/2, \lambda = 1/2)$ r.v.,s
$$V \sim Gamma(\alpha = n/2, \lambda = 1/2).$$
- Density function: $f(v) = \dfrac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}, v > 0.$

(α is the **shape parameter** and λ is the **scale parameter**)

Student's t Distribution

Definition. For *independent* r.v.'s Z and U where

- $Z \sim N(0,1)$
- $U \sim \chi^2_r$

the distribution of $T = Z/\sqrt{U/r}$ is the

t **distribution with** r **degrees of freedom**.

Properties

- The density function of T is

$$f(t) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi}\Gamma(r/2)}\left(1+\frac{t^2}{r}\right)^{-(r+1)/2}, \ -\infty < t < +\infty$$

- For what powers k does $E[T^k]$ converge/diverge?
- Does the moment generating function for T exist?

F Distribution

Definition. For *independent* r.v.'s U and V where

- $U \sim \chi^2_m$
- $V \sim \chi^2_n$

the distribution of $F = \dfrac{U/m}{V/n}$ is the

F **distribution with** m **and** n **degrees of freedom**.

(notation $F \sim F_{m,n}$)

Properties

- The density function of F is
$$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{n/2} w^{m/2-1} \left(1 + \frac{m}{n} w\right)^{-(m+n)/2},$$
with domain $w > 0$.
- $E[F] = E[U/m] \times E[n/V] = 1 \times n \times \frac{1}{n-2} = \frac{n}{n-2}$ (for $n > 2$).
- If $T \sim t_r$, then $T^2 \sim F_{1,r}$.

Outline

1. Distributions Derived from Normal Random Variables
   - χ^2, t, and F Distributions
   - Statistics from Normal Samples

Statistics from Normal Samples

Sample of size n from a Normal Distribution

- $X_1, \ldots, X_n$ *iid* $N(\mu, \sigma^2)$.
- **Sample Mean**: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- **Sample Variance**: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Properties of $\overline{X}$

- The moment generating function of $\overline{X}$ is

$$\begin{aligned} M_{\overline{X}}(t) &= E[e^{t\overline{X}}] = E[e^{\frac{t}{n}\sum_1^n X_i}] \\ &= \prod_1^n M_{X_i}(t/n) = \prod_{i=1}^n [e^{\mu(t/n) + \frac{\sigma^2}{2}(t/n)^2}] \\ &= e^{\mu t + \frac{\sigma^2/n}{2}t^2} \end{aligned}$$

i.e., $\overline{X} \sim N(\mu, \sigma^2/n)$.

Independence of $\overline{X}$ and S^2.

Theorem 6.3.A The random variable $\overline{X}$ and the vector of random variables

$$(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X}) \text{ are independent.}$$

Proof:

- Note that $\overline{X}$ and each of $X_i - \overline{X}$ are linear combinations of $X_1, \ldots, X_n$, i.e.,

$$\begin{aligned} \overline{X} &= \textstyle\sum_1^n a_i X_i = \mathbf{a}^T \mathbf{X} = U \text{ and} \\ X_k - \overline{X} &= \textstyle\sum_1^n b_i^{(k)} X_i = (\mathbf{b}^{(k)})^T \mathbf{X} = V_k \end{aligned}$$

where

$$\begin{aligned} \mathbf{X} &= (X_1, \ldots, X_n) \\ \mathbf{a} &= (a_1, \ldots, a_n) = (1/n, \ldots, 1/n) \\ \mathbf{b}^{(k)} &= (b_1^{(k)}, \ldots, b_n^{(k)}) \end{aligned}$$

$$\text{with } b_i^{(k)} = \begin{cases} 1 - \frac{1}{n}, & \textit{for} \quad i = k \\ -\frac{1}{n}, & \textit{for} \quad i \neq k \end{cases}$$

- U and $V_1, \ldots, V_n$ are jointly normal r.v.s
- U is uncorrelated (independent) of each V_k

Independence of $U = \mathbf{a}^T\mathbf{X}$ and $V = \mathbf{b}^T\mathbf{X}$

$$\begin{array}{rrcl} \text{where} & \mathbf{X} & = & (X_1, \ldots, X_n) \\ & \mathbf{a} & = & (a_1, \ldots, a_n) = (1/n, \ldots, 1/n) \\ & \mathbf{b} & = & (b_1, \ldots, b_n) \end{array}$$

with $b_i = b_i^{(k)} = \begin{cases} 1 - \frac{1}{n}, & \mathit{for} \quad i = k \\ -\frac{1}{n}, & \mathit{for} \quad i \neq k \end{cases}$ (so $V = V_k$)

Proof: Joint MGF of (U, V) factors into $M_U(s) \times M_V(t)$

$$\begin{aligned} M_{U,V}(s,t) &= E[e^{sU+tV}] = E[e^{s\mathbf{a}^T\mathbf{X}+t\mathbf{b}^T\mathbf{X}}] \\ &= E[e^{s[\sum_i a_iX_i]+t[\sum_i b_iX_i]}] = E[e^{\sum_i[(sa_i+tb_i)X_i]}] \\ &= E[e^{\sum_i[t_i^*X_i]}] \quad \text{with } t_i^* = (sa_i + tb_i) \\ &= \prod_{i=1}^n E[e^{[t_i^*X_i]}] = \prod_{i=1}^{n} M_{X_i}(t_i^*) \\ &= \prod_{i=1}^n e^{t_i^*\mu + \frac{\sigma^2}{2}(t_i^*)^2} = e^{\mu(\sum_{i=1}^n t_i^*) + \frac{\sigma^2}{2}\sum_{i=1}^n[(t_i^*)^2]} \\ &= e^{\mu(s\times 1 + t\times 0) + \sigma^2/2[\sum_{i=1}^n[s^2a_i^2 + t^2b_i^2 + 2sta_ib_i]} \\ &= [e^{s\mu + (\sigma^2/n)s^2/2}] \times [e^{t\times 0 + t^2\sigma^2/2\sum_{i=1}^n b_i^2}] \end{aligned}$$

$= [\text{mgf of } N(\mu, \sigma^2/n)] \times [\text{mgf of } N(0, \sigma^2 \cdot \sum_{i=1}^n b_i^2]$

Independence of $\overline{X}$ and S^2

Proof (continued):

- Independence of $U = \overline{X}$ and each $V_k = X_k - \overline{X}$ gives
  $\overline{X}$ and $S^2 = \sum_1^n V_i^2$ are independent.
- Random variables/vectors are independent if their joint moment generating function is the product of their individual moment generating functions:

$$\begin{aligned} M_{U,V_1,\ldots,V_n}(s,t_1,\ldots,t_n) &= E[e^{sU+t_1V_1+\cdots t_nV_n}] \\ &= M_U(s) \times M(t_1,\ldots,t_n). \end{aligned}$$

Theorem 6.3.B The distribution of $(n-1)S^2/\sigma^2$ is the Chi-Square distribution with $(n-1)$ degrees of freedom.

Proof:

- $\frac{1}{\sigma^2}\sum_1^n(X_k-\mu)^2 = \sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^n Z_i^2 \sim \chi_n^2.$
- $\frac{1}{\sigma^2}\sum_1^n(X_k-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n\left(X_i-\overline{X}+\overline{X}-\mu\right)^2$
  $= \frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\overline{X})^2 + \frac{n}{\sigma^2}\left(\overline{X}-\mu\right)^2$

Proof (continued):

Proof:

- $\frac{1}{\sigma^2}\sum_1^n (X_k-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X})^2 + \frac{n}{\sigma^2}\left(\overline{X}-\mu\right)^2$
- $\chi_n^2 = [\textit{distribution of } (nS^2/\sigma^2)] + \chi_1^2$
- By independence:
  mgf of $\chi_n^2 =$ mgf of $[\textit{distribution of } (nS^2/\sigma^2)] \times$ mgf of χ_1^2
  that is
  $$(1-2t)^{-n/2} = M_{nS^2/\sigma^2}(t) \times (1-2t)^{-1/2}$$
  So
  $$M_{nS^2/\sigma^2}(t) = (1-2t)^{-(n-1)/2},$$
  $$\Longrightarrow nS^2/\sigma^2 \sim \chi_{n-1}^2.$$

Corollary 6.3.B For a $\overline{X}$ and S^2 from a Normal sample of size n,

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$