4. Densities in general:

   a. General in sense of j not necessarily either 1 or n

      i. Still require continuity, independence, same distribution.

   b. Pick j and value x where you want to evaluate the probability density function of $X_{(j)}$.

   c. $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1 - F_X(x))^{n-j}$, because:

      i. Define (U_1, U_2, U_3) be counts of observations in various ranges.

      - Pick Δ defining the width of a small interval around x.
      - U_1 is the number of observations in $(-\infty, x - \Delta/2]$
      - U_2 is the number of observations in $(x - \Delta/2, x + \Delta/2)$
      - U_3 is the number of observations in $[x + \Delta/2, \infty)$

      ii. (U_1, U_2, U_3) is multinomial with cell probabilities depending on F_X

      - $(F_X(x - \Delta/2), F_X(x + \Delta/2) - F_X(x - \Delta/2), 1 - F_X(x + \Delta/2))$.

      iii. For Δ small enough, $\mathrm{P}\left(X_{(j)} \in (x - \Delta/2, x + \Delta/2)\right) \approx \mathrm{P}\left(U_1 = j - 1, U_2 = 1, U_3 = n - j\right)$

- Only approximate, because there's a probability that two or more observations sit in the center bin.
- This probability is less than $C\Delta^2$ for some C.
- It will disappear in the limit.

iv. $\mathsf{P}(U_1 = j-1, U_2 = 1, U_3 = n-j) = \frac{n!}{(j-1)!1!(n-j)!} F_X(x-\Delta/2)^{j-1}(F_X(x+\Delta/2) - F_X(x-\Delta/2))(1-F_X(x+\Delta/2))^{n-j}$

v. For Δ small enough, $\mathsf{P}\left(X_{(j)} \in (x-\Delta/2, x+\Delta/2)\right) \approx \frac{n!}{(j-1)!1!(n-j)!} F_X(x-\Delta/2)^{j-1} f_X(x)\Delta(1-F_X(x+\Delta/2))^{n-j}$.

- From approximating the distribution function difference as the interval length times the probability density function in the middle.

vi. $f_{X_{(j)}}(x) = \lim_{\Delta\to 0} \mathsf{P}\left(X_{(j)} \in (x-\Delta/2, x+\Delta/2)\right)/\Delta = \frac{n!}{(j-1)!1!(n-j)!} F_X(x)^{j-1} f_X(x)(1-F_X(x))^{n-j}$.

d. Special case of median for odd sample size:

i. For $n = 2r+1$,

ii. then $f_{X_{(r+1)}}(x) = \frac{n!}{r!r!} F_X(x)^r f_X(x)(1-F_X(x))^r$.

5. Examples:

a. $X_j \sim$ Expon with rate λ.

i. $F_X(x) = 1 - \exp(-\lambda x)$.

ii. $1 - F_X(x) = \exp(-\lambda x)$.

iii. $F_{X_{(1)}}(x) = 1 - \exp(-\lambda x)^n = 1 - (\exp(-\lambda x))^n = 1 - (\exp(-n\lambda x))$.

iv. Hence $X_{(1)} \sim$ Expon with rate $n\lambda$

v. $F_{X_{(n)}}(x) = (1 - \exp(-\lambda x))^n$: no significant simplification.

b. $X_j \sim$ Beta$(1, 1) =$ Unif$(0, 1)$.

i. $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} u^{j-1}(1-u)^{n-j}$

ii. $X_{(j)} \sim$ Beta$(j, n-j+1)$.

WMS: 7.1-7.2

K. Distributions Derived from the Normal

1. Distribution of the Sample Mean $\bar{X} = \sum_{i=1}^{n} X_i/n$.

a. Let $\mu =$ E(X_i), $\sigma^2 =$ V(X_i).

b. Moments Without Assuming Independence:

i. E$\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}$ E$(X_i) = n\mu$.

ii. E$\left(\bar{X}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \mu$.

c. Moments Assuming Independence: V$\left(\bar{X}\right) =$ V$\left(\sum_{i=1}^{n} X_i\right)\left(\frac{1}{n}\right)^2 = \sigma^2/n$, because

i. $\mathsf{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{V}\left(X_i\right) = n\sigma^2$.

ii. Moment calculations do not require normality.

d. Shape (assuming independence, normality):

i. Sum of two bivariate normals is again normal

ii. Two independent normals is a trivial case of bivariate normal.

iii. Inductively, $\sum_{j=1}^{n} X_j$ normal

iv. Hence $\bar{X} = \sum_{j=1}^{n} X_j/n$ is normal.

2. Distribution of Sum of Squares

a. Sum of Squares from mean can be written as the sum of squared independent random variables.

b. Let $Q_n = \sum_{j=1}^{n}(X_j - \bar{X}_n)^2$, $\bar{X}_n = \sum_{j=1}^{n} X_j/n$

i. Evaluating, $Q_2 = (X_1 - (X_1 + X_2)/2)^2 + (X_2 - (X_1 + X_2)/2)^2 = 2((X_1 - X_2)/2)^2 = (X_1 - X_2)^2/2$.

ii. Then $\bar{X}_{n-1} = \sum_{j=1}^{n-1} X_j/(n-1)$

iii. Above example shows $Q_2 = (X_2 - X_1)^2(1 - 1/2) = (X_2 - \bar{X}_1)^2(1 - 1/2)$.

c. Express in terms of quantities with last omitted:

$$Q_n = \sum_{j=1}^{n}(X_j - \bar{X}_n)^2 = \sum_{j=1}^{n-1}(X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2 .$$

i. Note $\bar{X}_n = \frac{1}{n}X_n + \frac{n-1}{n}\bar{X}_{n-1}$.

ii. So $Q_n = \sum_{j=1}^{n-1}(X_j - \bar{X}_n)^2 + (X_n - \bar{X}_n)^2 = \sum_{j=1}^{n-1}\left(X_j - \frac{1}{n}X_n - \frac{n-1}{n}\bar{X}_{n-1}\right)^2 + (\frac{n-1}{n})^2(X_n - \bar{X}_{n-1})^2$

d. Claim: $Q_n = \sum_{j=2}^{n}(X_j - \bar{X}_{j-1})^2(1 - 1/j)$.

i. Works for $n = 2$.

ii. Suppose $Q_{n-1} = \sum_{j=2}^{n-1}(X_j - \bar{X}_{j-1})^2(1 - 1/j)$.

iii. Square out X_n and the rest:

$$Q_n = \left(\left(1 - \frac{1}{n}\right)X_n - \frac{n-1}{n}\bar{X}_{n-1}\right)^2 + \sum_{j=1}^{n-1}(\tfrac{1}{n}X_n - X_j + \tfrac{n-1}{n}\bar{X}_{n-1})^2$$

iv. Collect terms:

$$Q_n = \left((1 - \tfrac{1}{n})^2 + \tfrac{n-1}{n^2}\right)X_n^2 + 2X_n\left[-n(1 - \tfrac{1}{n})^2\bar{X}_{n-1} + \tfrac{(n-1)^2}{n^2}\bar{X}_{n-1} - \tfrac{n-1}{n}\bar{X}_{n-1}\right] + \left[\tfrac{(n-1)^2}{n^2}\bar{X}_{n-1}^2 + \sum_{j=1}^{n-1}(X_j - \frac{n-1}{n}\bar{X}_{n-1})^2\right]$$

v. Simplify:

$$Q_n = \left(1 - \frac{1}{n}\right)[X_n^2 - 2X_n\bar{X}_{n-1}] + \frac{(n-1)^2}{n^2}\bar{X}_{n-1}^2 + Q_{n-1} + n((n-1)/n - 1)^2\bar{X}_{n-1}^2$$

vi. Simplify some more: $Q_n = (1 - 1/n)[X_n^2 - 2X_n\bar{X}_{n-1} + \bar{X}_{n-1}^2] + Q_{n-1}$.

vii. Identify the square: $(1 - 1/n)[X_n - \bar{X}_{n-1}]^2 + Q_{n-1}$.

viii. By induction, holds for all n.

e. Before squaring, each summand has expectation 0.

f. Summands are independent:

i. For $j > i > 1$,

- Let $a_k = \begin{cases} 0 & \text{if } k > j \\ 1 & \text{if } k = j \\ -1/(j-1) & \text{if } k < j \end{cases}$, $b_k = \begin{cases} 0 & \text{if } k > i \\ 1 & \text{if } k = i \\ -1/(i-1) & \text{if } k < i \end{cases}$.

ii. a and b quantities are multiples of rows of

$$\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ -1/2 & -1/2 & 1 & 0 & \cdots & \cdots & 0 \\ -1/3 & -1/3 & -1/3 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & & & & \\ -1/(n-1) & -1/(n-1) & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 1 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

iii. Cov $\left(X_i - \bar{X}_{i=1}, X_j - \bar{X}_{j-1}\right) = \sum_{k=1}^{n} b_k a_k \sigma^2 = \left(1 - \frac{i-1}{i-1}\right) \sigma^2 = 0$.

iv. Hence these are independent.

g. Before squaring, each summand has variance σ^2.

i. V $\left(\sqrt{\frac{n-1}{n}}(X_n - \bar{X}_{n-1})\right) = \frac{n-1}{n}\left(1 + \frac{1}{n-1}\right)\sigma^2 = \sigma^2$.

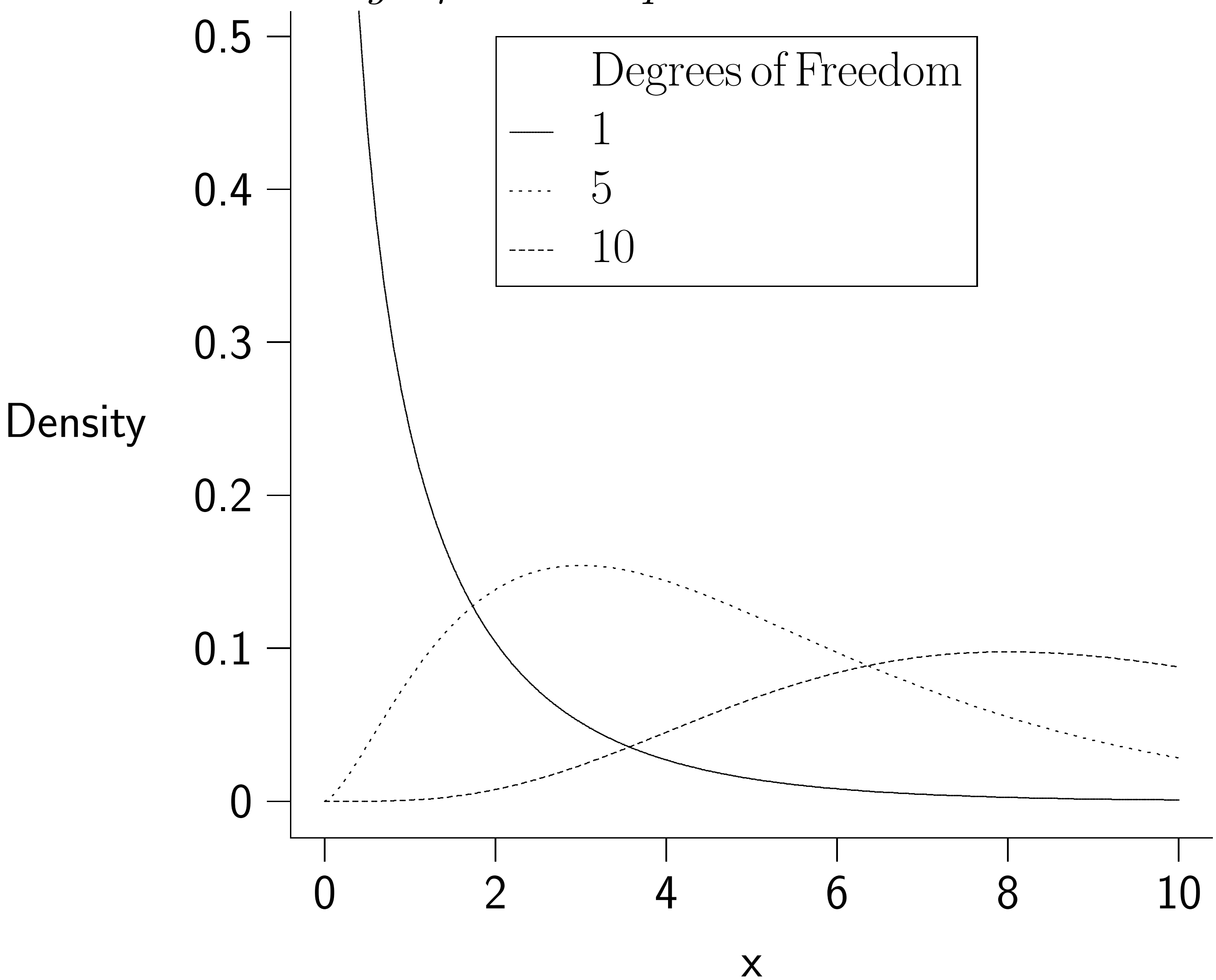
h. Hence Q_n/σ^2 has same distribution as sum of $n-1$ independent squared normals: χ^2_{n-1}.

i. As before, χ^2_{n-1} is a special case of gamma.

i. $n-1$ is called the *degrees of freedom*.

i. Distribution depends on degrees of freedom. See Fig. 43.

Fig. 43: Chi-Square Densities



j. Furthermore, $\bar{X}$ and Q_n are independent.

i. Because for any $j \leq n$, Cov $\left(\bar{X}, X_j - \bar{X}_{j-1}\right) = \sum_{k=1}^{n}(1/n)a_j\sigma^2 = 0$.

k. R calculates distributional quantities for $W \sim \chi_k^2$

i. Calculate probabilities P $(W \leq w)$ using `pchisq(w,k)`.

ii. Calculate quantiles using `qchisq(q,k)` for $q \in (0,1)$.