Important

Random samples and iid variables

Distribution of Maximum and Minimum

Let $X_1, \dots, X_n$ be a random sample with common cdf $F_{X_1}(x) = P(X_1 \leq x)$

Let $X_{(n)} = \max\{X_1, \dots, X_n\}$ and $X_{(1)} = \min\{X_1, \dots, X_n\}$

Important results:

1. $F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = [F_{X_1}(x)]^n$, for $x \in \mathbb{R}$

2. $F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - [1 - F_{X_1}(x)]^n$, for $x \in \mathbb{R}$

3. If the population cdf $F_{X_1}(x) = P(X_1 \leq x)$ is continuous with pdf $f_{X_1}(x) = \frac{dF_{X_1}(x)}{dx}$, then $X_{(n)}$ and $X_{(1)}$ both have pdfs given by

$$f_{X_{(n)}}(x) = n f_{X_1}(x)[F_{X_1}(x)]^{n-1}, \qquad f_{X_{(1)}}(x) = n f_{X_1}(x)[1 - F_{X_1}(x)]^{n-1}$$

Proofs: (These are proofs that are useful to remember.)

Random samples and iid variables

Order statistics

- *Definition:* The **order statistics** for a sample $X_1, \ldots, X_n$ are the values in ascending order denoted as

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

- Primarily interested in iid $X_1, \ldots, X_n$ having a continuous distribution

- For random samples we may be interested in
  1. the distribution of a single order statistic $X_{(i)}$
  2. the distribution of two or more order statistics $(X_{(i)}, X_{(j)})$
  3. function of two or more order statistics

     e.g., range $R = X_{(n)} - X_{(1)}$

$X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathrm{Uni}(\theta_1, \theta_2)$

θ_1, θ_2 are Unknown parameters.

$\underline{(X_{(1)}, X_{(n)})}$ is the $\underline{\text{MLE}}$ for (θ_1, θ_2). STAT5430

$\mathbb{E}X_{(1)}$, $\mathbb{E}X_{(n)}$ to say Something about θ_1 and θ_2.

$\mathbb{E}[U(T)] = 0 \Rightarrow U = 0$ a.s. $\Rightarrow$ T is UMVUE

$T = h(X_{(1)}, X_{(n)})$

- order statistics are a type of (discontinuous) transformation of $X_1, \ldots, X_n$

Random samples and iid variables

Distribution of kth order statistic

Result 1: If $X_1, \ldots, X_n$ are a random sample with common cdf $F_{X_1}(x)$, then the cdf of the kth order statistic (given some $k = 1, \ldots, n$) is given by

$$\underline{F_{X_{(k)}}(x)} = P(X_{(k)} \leq x) = P(\text{at least } k\ X_i\text{'s} \leq x) = \boxed{\sum_{j=k}^{n} \binom{n}{j} [F_{X_1}(x)]^j [1 - F_{X_1}(x)]^{n-j}}$$

Proof:

Result 2 (pdf in continuous case): If $X_1, \ldots, X_n$ are a random sample with common continuous cdf $F_{X_1}(x)$ and pdf $f_{X_1}(x)$, the pdf of the kth order statistic is

result

$$f_{X_{(k)}}(x) = \frac{dF_{X_{(k)}}(x)}{dx} = \boxed{\frac{n!}{(k-1)!(n-k)!}} f_{X_1}(x) [F_{X_1}(x)]^{k-1} \underbrace{[1 - F_{X_1}(x)]^{n-k}}_{X_i\text{'s are greather } x}$$

- Heuristic argument for the form of the pdf $f_{X_{(k)}}(x)$:

  $\underbrace{k - 1 \text{ observations} \leq x}_{X_i \leq x}$; $\overbrace{1 \text{ observation in } (x, x + dx)}^{\mathbb{P}(x < X_{(k)} < x + dx)}$; $\underbrace{n - k \text{ observations} > x}$

  x

  $X_{(k)}$

  $\binom{n-1}{k-1}$

  k-1 observations are less than x

  n-k observations are greater than x

- A formal proof uses derivative of cdf + algebra (see next slide)

Note: in the discrete case, the pmf of $X_{(k)}$ is obtained as

$$f_{X_{(k)}}(x) = \underbrace{P(X_{(k)} = x)}_{\text{pmf}} = \underbrace{P(X_{(k)} \leq x) - P(X_{(k)} < x)} = F_{X_{(k)}}(x) - \lim_{y \uparrow x} F_{X_{(k)}}(y)$$

Random samples and iid variables

Distribution of kth order statistic (cont'd)

Proof of the pdf of the kth order statistic:

(Skip or read as you wish!)

Recall the cdf is $F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} [F_{X_1}(x)]^j [1 - F_{X_1}(x)]^{n-j}$

$$
\begin{aligned}
f_{X_{(k)}}(x) = \frac{dF_{X_{(k)}}(x)}{dx} &= \sum_{j=k}^{n} \binom{n}{j} j f_{X_1}(x) [F_{X_1}(x)]^{j-1} [1 - F_{X_1}(x)]^{n-j} \\
&\quad - \sum_{j=k}^{n} \binom{n}{j} (n-j) f_{X_1}(x) [F_{X_1}(x)]^{j} [1 - F_{X_1}(x)]^{n-j-1} \\
&= \binom{n}{k} k f_{X_1}(x) [F_{X_1}(x)]^{k-1} [1 - F_{X_1}(x)]^{n-k} + R_1 + R_2
\end{aligned}
$$

$$
\begin{aligned}
R_1 &\equiv \sum_{j=k+1}^{n} \binom{n}{j} j f_{X_1}(x) [F_{X_1}(x)]^{j-1} [1 - F_{X_1}(x)]^{n-j} \\
&= \sum_{z=k}^{n-1} \binom{n}{z+1} (z+1) f_{X_1}(x) [F_{X_1}(x)]^{z} [1 - F_{X_1}(x)]^{n-z-1} \quad (z = j-1) \\
&= \sum_{z=k}^{n-1} \binom{n}{z} (n-z) f_{X_1}(x) [F_{X_1}(x)]^{z} [1 - F_{X_1}(x)]^{n-z-1}
\end{aligned}
$$

using $\binom{n}{z+1}(z+1) = \frac{n!}{(n-z-1)!z!} = \frac{n!}{(n-z)!z!}(n-z) = \binom{n}{z}(n-z)$

$$
\begin{aligned}
R_2 &\equiv -\sum_{j=k}^{n} \binom{n}{j} (n-j) f_{X_1}(x) [F_{X_1}(x)]^{j} [1 - F_{X_1}(x)]^{n-j-1} \\
&= -\sum_{j=k}^{n-1} \binom{n}{j} (n-j) f_{X_1}(x) [F_{X_1}(x)]^{j} [1 - F_{X_1}(x)]^{n-j-1}
\end{aligned}
$$

Hence $R_1 + R_2 = 0$

Random samples and iid variables

Joint distribution of other order statistics

We'll just consider the continuous case and state distributions in terms of pdfs

For random sample $X_1, \ldots, X_n$ with common continuous cdf $F_{X_1}(x)$ and pdf $f_{X_1}(x)$,

**

1. joint pdf of two order statistics $(X_{(i)}, X_{(j)})$ with $i < j$

   n observations

   $\underbrace{\quad\quad}_{}$ (i-1) $\quad$ $X_{(i)}$ is here $\quad$ $\underbrace{\quad\quad}_{}$ j-i-1 $\quad$ $X_{(j)}$ is here $\quad$ $\underbrace{\quad\quad}_{}$ n-j obs.

   $$f_{X_{(i)},X_{(j)}}(u, v) =$$

   $$\frac{n!}{(i-1)!(j-i-1)!(n-j!)} f_{X_1}(u) f_{X_1}(v) [F_{X_1}(u)]^{i-1} [F_{X_1}(v) - F_{X_1}(u)]^{j-i-1} [1 - F_{X_1}(v)]^{n-j}$$

   for $u < v$

*

2. joint pdf for all order statistics

   $$f_{X_1,X_2,\ldots,X_n}(u_1,u_2,\ldots,u_n) \overset{\text{independent}}{=} f_{X_1}(u_1) \cdots f_{X_n}(u_n) \overset{\text{identically dist.}}{=} f_{X_1}(u_1) \cdots f_{X_1}(u_n)$$

   $$f_{X_{(1)},\cdots,X_{(n)}}(u_1, \cdots, u_n) = n! f_{X_1}(u_1) f_{X_1}(u_2) \cdots f_{X_1}(u_n) \longrightarrow X_1 \cdots X_n \text{ are random sample (i.i.d)}$$

   n! counts the # of ways to order $X_1, \ldots, X_n$

   for $u_1 < \cdots < u_n$

Random samples and iid variables

Order statistics: example

Example: Let $X_1, \ldots, X_n$ be a sample random from Uniform$(0,1)$

Find the pdf of $X_{(k)}$.

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!}\left[F_{X_1}(x)\right]^{k-1} \underbrace{f_{X_1}(x)}_{(1)} \underbrace{\left[1-F_{X_1}(x)\right]^{n-k}}_{(1-x)^{n-k}}$$

$$= \frac{n!}{(k-1)!\,(n-k)!}\left(x^{k-1}\right)(1)(1-x)^{n-k} \qquad 0<x<1$$

$$\underbrace{F_{X_1}(x)} \overset{def}{=} \mathbb{P}(X_1 \le x) = \int_0^x 1\,dy = x$$

$X_1 \sim Uni(0,1)$

$f_{X_1}(y)=1$ for $0<y<1$

$$1 - F_{X_1}(x) = 1-x$$

$$f_{X_{(k)}}(x) = \begin{cases} \dfrac{n!}{\underbrace{(k-1)!}\,\underbrace{(n-k)!}} x^{k-1} (1-x)^{(n-k+1)-1}, & 0<x<1 \\ 0, & ow \end{cases}$$

$$X_{(k)} \sim \text{Beta}(k, n-k+1)$$

$$\text{Beta function}(k, n-k+1) = \frac{\Gamma(\cancel{k}+n-\cancel{k}+1)}{\Gamma(k)\,\Gamma(n-k+1)}$$

$$\boxed{\begin{array}{l} \text{If } X_1, \ldots, X_n \overset{i.i.d}{\sim} Uni(0,1) \\ \Rightarrow X_{(k)} \sim \text{Beta}(k, n-k+1) \text{ dist.} \end{array}}$$

Sampling from the Normal Distribution

Review of random samples

- Recall that $X_1, \ldots, X_n$ iid with pdf $f_X(x)$ means

$$\longrightarrow f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i)$$

- A function $T = T(X_1, \ldots, X_n)$ of the random variables is a **statistic**

- Previous results for random samples with mean $\mathrm{E}X_1$ and variance $\mathrm{Var}(X_1)$:

$X_1, X_2, \ldots, X_n$ i.i.d $\mathbb{E}X_i = \mu$, $Var\, X_i = \sigma^2$

The sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ has

$$\mathrm{E}(\bar{X}_n) = \mathrm{E}X_1 = \mu \qquad \mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X_1)}{n} = \frac{\sigma^2}{n}$$

The sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ has

$$\mathrm{E}(S^2) = \mathrm{Var}(X_1) = \sigma^2 \qquad \mathrm{Var}(S^2) = \frac{1}{n}\left(\mathrm{E}(X_1 - \mu)^4 - \frac{n-3}{n-1}\sigma^4\right)$$

- If the distribution of the X_i's is normal (i.e., $X_1, \ldots, X_n$ iid $\boxed{\sim N(\mu, \sigma^2)}$), then we can derive the $\boxed{\textit{exact}}$ distribution of $\bar{X}_n, S^2$ *for any* n

"normal Sampling theory"

Sampling from the Normal Distribution

Preliminary results/facts

1. If $\underbrace{Z \sim N(0,1)}$ then $Z^2 \sim \chi^2_1$.

2. If $X \sim N(\mu, \sigma^2)$ then $\underbrace{(X-\mu)^2/\sigma^2} \sim \chi^2_1$

$$X \sim N(\mu, \sigma^2) \Rightarrow \underbrace{\frac{X-\mu}{\sigma}}_{Z} \sim N(0,1) \xrightarrow{\text{Result 1}} \underbrace{\left(\frac{X-\mu}{\sigma}\right)^2}_{= Z^2} \sim \chi^2_1$$

**
3. If $Y_1, \ldots, Y_n$ are independent r.v.s where $Y_i \sim \chi^2_{\nu_i}$, then $Y = \sum_{i=1}^n Y_i \sim \chi^2_{\sum_{i=1}^n \nu_i}$

Proof: We've basically seen this already: use mgf technique for sums

$Y_i \sim \chi^2_{\nu_i}$

$$M_Y(t) = \mathrm{E}e^{tY} = \mathrm{E}e^{t\sum_{i=1}^n Y_i} = \mathrm{E}\prod_{i=1}^n e^{tY_i} = \prod_{i=1}^n \mathrm{E}e^{tY_i} = \prod_{i=1}^n M_{Y_i}(t)$$

def of $M_Y(t)$ def of Y $e^{x+y} = e^x \cdot e^y$ Y_i's are independent def of MGF

$$= \prod_{i=1}^n (1-2t)^{-\nu_i} = (1-2t)^{-\sum_{i=1}^n \nu_i}$$

Provided $t < 1/2$

MGF of $Y \Rightarrow Y \sim \chi^2_{\sum_{i=1}^n \nu_i}$

We will see next week, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

4. If $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ then $\underbrace{\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2} \sim \chi^2_n$

$$\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 = \underbrace{\underbrace{\left(\frac{X_1-\mu}{\sigma}\right)^2}_{\chi^2_1} + \underbrace{\left(\frac{X_2-\mu}{\sigma}\right)^2}_{\chi^2_1} + \cdots + \underbrace{\left(\frac{X_n-\mu}{\sigma}\right)^2}_{\chi^2_1}} \Rightarrow \chi^2_{\sum_{i=1}^n \nu_i} = \chi^2_n$$

+ X_i's are independent