

MATH 556: MATHEMATICAL STATISTICS I

BASIC EXCHANGEABILITY CONSTRUCTIONS

An infinite sequence of random variable $X_1, X_2, \dots, X_n, \dots$ is *exchangeable* (or *infinitely exchangeable*) if, for any $n \geq 1$ and sets $A_1, A_2, \dots, A_n \subseteq \mathbb{R}$ we have that

$$P_{X_1, \dots, X_n} \left[\bigcap_{j=1}^n (X_j \in A_j) \right] = P_{X_{\tau(1)}, \dots, X_{\tau(n)}} \left[\bigcap_{j=1}^n (X_{\tau(j)} \in A_j) \right]$$

for all permutations $(\tau(1), \dots, \tau(n))$ of the labels $(1, \dots, n)$. In terms of cdfs, we can express this as that for all $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_{\tau(1)}, \dots, X_{\tau(n)}}(x_1, \dots, x_n).$$

We have the following characterization: the infinite sequence $X_1, X_2, \dots, X_n, \dots$ is exchangeable if and only if the representation

$$P_{X_1, \dots, X_n} \left[\bigcap_{j=1}^n (X_j \in A_j) \right] = \int \left\{ \prod_{j=1}^n P_{X_j|T}[X_j \in A_j|T = t] \right\} dF_T(t)$$

holds for some other random variable T with distribution F_T . That is, the sequence is exchangeable if and only if elements in the sequence are conditionally independent given T , for some T with distribution F_T . In fact, T is a random variable formed as some function of (X_1, \dots, X_n) in the limiting case as $n \rightarrow \infty$.

The representation also indicates that we can construct exchangeable random variables by following the construction

$$\begin{aligned} T &\sim f_T(t) \\ X_1, \dots, X_n &\sim f_{X|T}(x|t) \quad \text{independent} \end{aligned}$$

EXAMPLE: Suppose $T \sim \text{Uniform}(0, 1)$, and $X_1, \dots, X_n|T = t \sim \text{Bernoulli}(t)$ independently. Then

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_0^1 \prod_{j=1}^n f_{X_j|T}(x_j|t) f_T(t) dt = \int_0^1 t^s (1-t)^{n-s} dt = \frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)}$$

where $s = \sum_{j=1}^n x_j$, for $s = 0, 1, \dots, n$, where the support of the joint pmf is the set $\{0, 1\}^n$ of binary vectors of length n . The integral is analytically tractable as the integrand is proportional to a *Beta*($s+1, n-s+1$) pdf. Note that in this construction, the quantity s is associated with a corresponding random variable

$$S = \sum_{i=1}^n X_i$$

which we can consider a *summary statistic*, and notice that the event $S = s$ corresponds to

$$\binom{n}{s}$$

individual sequences of x values which all have the same joint probability: this demonstrates exchangeability. Thus

$$f_S(s) = \binom{n}{s} \frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)} = \frac{n!}{s!(n-s)!} \frac{s!(n-s)!}{(n+1)!} = \frac{1}{n+1} \quad s = 0, 1, \dots, n.$$

and zero otherwise.

```

n<-10
s<-0:n
fs<-choose(n,s)*gamma(s+1)*gamma(n-s+1)/gamma(n+2)
fs

+ [1] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909
+ [7] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909

sum(fs)

+ [1] 1

sim.exch01<-function(nv){ #Sample the exchangeable binary variables.
  Tv<-runif(1)
  Xv<-rbinom(nv,1,Tv)
}
svals<-replicate(10000,sum(sim.exch01(n))) #10000 replicate draws of S
table(svals)/10000

+ svals
+ 0 1 2 3 4 5 6 7 8 9 10
+ 0.0891 0.0898 0.0912 0.0930 0.0921 0.0904 0.0852 0.0936 0.0893 0.0926 0.0937

```

EXAMPLE: Suppose $T \sim Normal(0, 1)$, and $X_1, \dots, X_n | T = t \sim Normal(t, 1)$ independently. Then

$$\begin{aligned}
f_{X_1, \dots, X_n}(x_1, \dots, x_n) &= \int_{-\infty}^{\infty} \prod_{j=1}^n f_{X_j | T}(x_j | t) f_T(t) dt \\
&= \int_{-\infty}^{\infty} \prod_{j=1}^n \left\{ \left(\frac{1}{2\pi} \right)^{1/2} \exp \left\{ -\frac{1}{2} (x_j - t)^2 \right\} \right\} \left(\frac{1}{2\pi} \right)^{1/2} \exp \left\{ -\frac{1}{2} t^2 \right\} dt \\
&= \left(\frac{1}{2\pi} \right)^{(n+1)/2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n (x_j - t)^2 + t^2 \right] \right\} dt.
\end{aligned}$$

Now, using the completing the square formula

$$A(t - a)^2 + B(t - b)^2 = (A + B) \left(t - \frac{Aa + Bb}{A + B} \right)^2 + \frac{AB}{A + B} (a - b)^2$$

we have

$$\sum_{j=1}^n (x_j - t)^2 + t^2 = \sum_{j=1}^n (x_j - \bar{x})^2 + (n + 1) \left(t - \frac{n\bar{x}}{n + 1} \right)^2 + \frac{n}{n + 1} \bar{x}^2$$

so therefore, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n (x_j - t)^2 + t^2 \right] \right\} dt &= \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n (x_j - \bar{x})^2 + \frac{n}{n + 1} \bar{x}^2 \right] \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(n + 1)}{2} \left(t - \frac{n\bar{x}}{n + 1} \right)^2 \right\} dt \\
&= \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n (x_j - \bar{x})^2 + \frac{n}{n + 1} \bar{x}^2 \right] \right\} \sqrt{\frac{2\pi}{n + 1}}
\end{aligned}$$

as the integrand is proportional to a Normal pdf. Thus for $(x_1, \dots, x_n) \in \mathbb{R}^n$,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \left(\frac{1}{2\pi} \right)^{n/2} \sqrt{\frac{1}{n + 1}} \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^n (x_j - \bar{x})^2 + \frac{n}{n + 1} \bar{x}^2 \right] \right\}$$

which also relies only upon the summary statistics

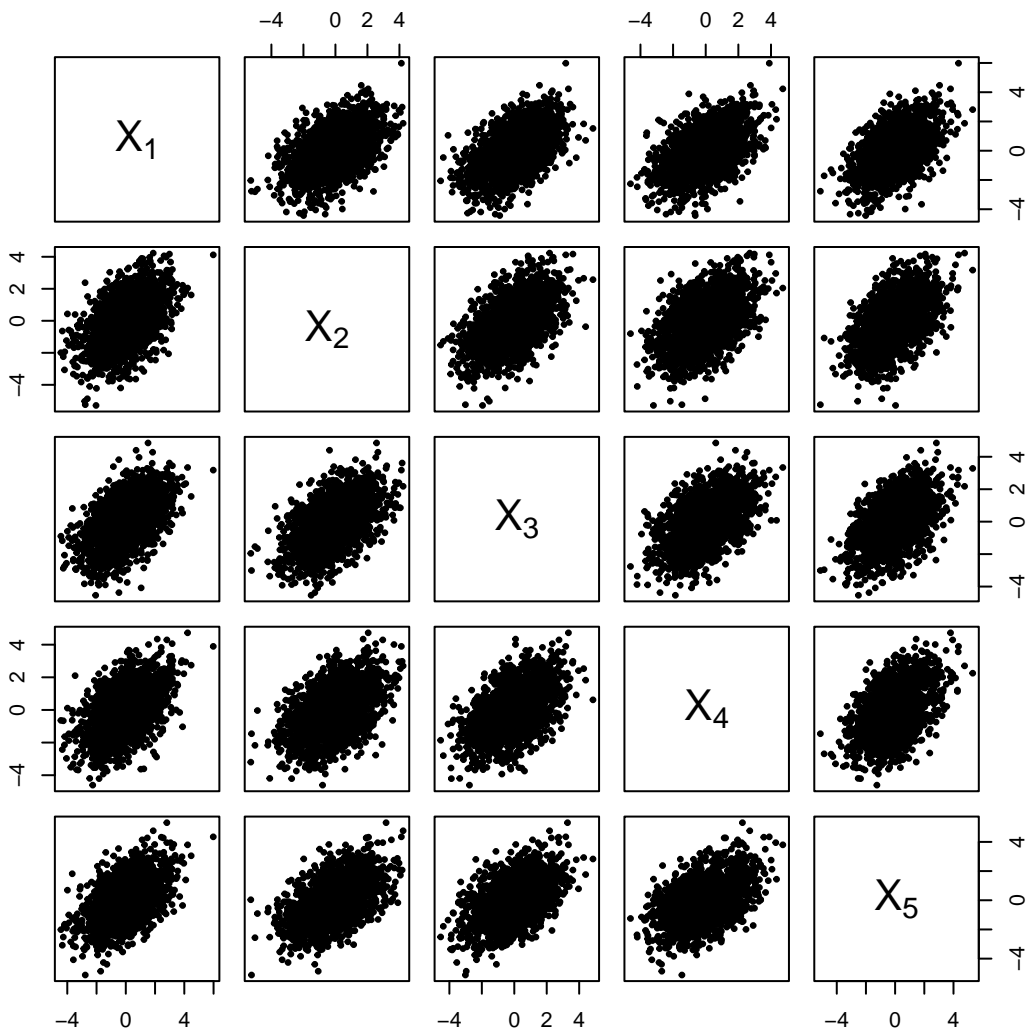
$$S_1 = \bar{x} \quad S_2 = \sum_{j=1}^n (x_j - \bar{x})^2$$

and so we observe exchangeability.

```
n<-10
sim.exch02<-function(nv){ #Sample the exchangeable variables.
  Tv<-rnorm(1)
  Xv<-rnorm(nv,Tv,1)
}
Xmat<-t(replicate(2000,sim.exch02(n))) #10000 replicate draws of S
dim(Xmat)

+ [1] 2000 10

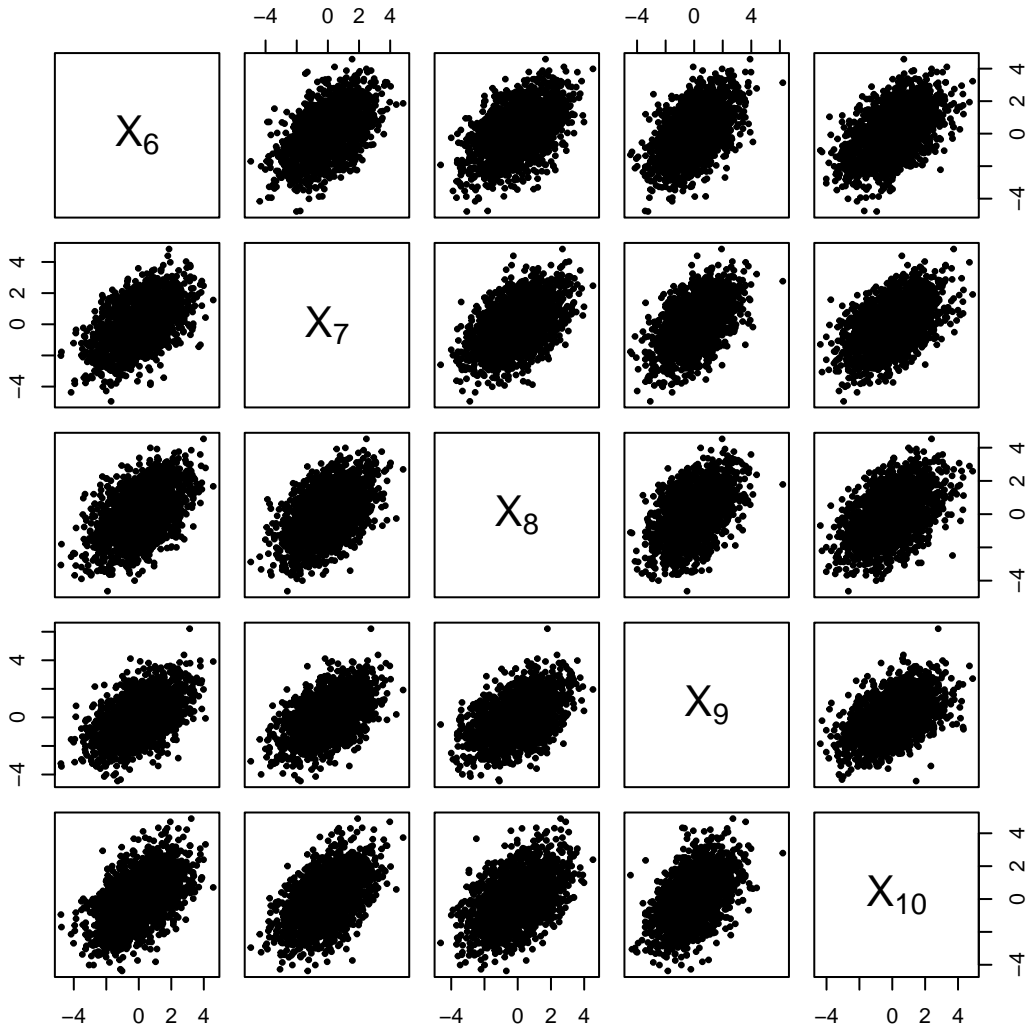
par(pty='s')
pairs(Xmat[,1:5],pch=19,cex=0.5,
      labels=c(expression(X[1]),expression(X[2]),expression(X[3]),
                expression(X[4]),expression(X[5])))
```



```

pairs(Xmat[,6:10],pch=19,cex=0.5,
      labels=c(expression(X[6]),expression(X[7]),expression(X[8]),
                expression(X[9]),expression(X[10])))

```



We have that for $j = 1, \dots, n$,

$$\mathbb{E}_{X_j}[X_j] = 0 \quad \text{Var}_{X_j}[X_j] = 2$$

```

apply(Xmat,2,mean)

```

```

+ [1] -0.027053489 -0.015066664 -0.001263267 -0.021293136 -0.019866554
+ [6]  0.031588502 -0.037480639 -0.024370383  0.005652952  0.018982775

```

```

apply(Xmat,2,var)

```

```

+ [1] 1.978509 2.081357 1.890184 1.986101 1.962283 1.940234 1.965412 1.938499
+ [9] 1.910753 1.992981

```

Also, for the covariances, using iterated expectation we have

$$\text{Cov}_{X_j, X_k}[X_j, X_k] \equiv \mathbb{E}_{X_j, X_k}[X_j X_k] = \mathbb{E}_T [\mathbb{E}_{X_j, X_k|T}[X_j X_k|T]] = \mathbb{E}_T [\mathbb{E}_{X_j|T}[X_j|T] \mathbb{E}_{X_k|T}[X_k|T]]$$

as X_j and X_k have expectation zero, and are conditionally independent given T . Thus, as $\mathbb{E}_{X_j|T}[X_j|T] = T$ for each j , we have

$$\text{Cov}_{X_j, X_k}[X_j, X_k] = \mathbb{E}_T[T^2] = 1$$

and hence

$$\text{Corr}_{X_j, X_k}[X_j, X_k] = \frac{\text{Cov}_{X_j, X_k}[X_j, X_k]}{\sqrt{\text{Var}_{X_j}[X_j]\text{Var}_{X_k}[X_k]}} = \frac{1}{2}.$$

```
round(cor(Xmat), 3)
```

```
+      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
+ [1,] 1.000 0.494 0.508 0.500 0.513 0.500 0.493 0.496 0.521 0.506
+ [2,] 0.494 1.000 0.510 0.480 0.504 0.488 0.484 0.480 0.481 0.487
+ [3,] 0.508 0.510 1.000 0.490 0.499 0.492 0.503 0.490 0.505 0.472
+ [4,] 0.500 0.480 0.490 1.000 0.497 0.505 0.500 0.483 0.477 0.514
+ [5,] 0.513 0.504 0.499 0.497 1.000 0.506 0.480 0.476 0.484 0.498
+ [6,] 0.500 0.488 0.492 0.505 0.506 1.000 0.519 0.488 0.491 0.501
+ [7,] 0.493 0.484 0.503 0.500 0.480 0.519 1.000 0.480 0.495 0.502
+ [8,] 0.496 0.480 0.490 0.483 0.476 0.488 0.480 1.000 0.477 0.487
+ [9,] 0.521 0.481 0.505 0.477 0.484 0.491 0.495 0.477 1.000 0.480
+ [10,] 0.506 0.487 0.472 0.514 0.498 0.501 0.502 0.487 0.480 1.000
```