Chapter 4
Multiple Random Variables
Order Statistics

Outline

Basic Definitions

The order statistics

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

is the increasing ordered arrangement of the sample

$$X_1, X_2, \ldots X_n.$$

The sample range $R = X_{(n)} - X_{(1)}$. The sample median

$$M = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd.} \\ X_{(n/2)} + X_{(n/2+1)} & \text{if } n \text{ is even.} \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be independent continuous random variables with common distribution function F_X and density f_X. Thus,

$$P\{X_i = X_j \text{ for some } i \neq j\} = 0.$$

Marginal Distribution Functions

Fix a value x and consider the Bernoulli trials

$$Y_1 = I_{(-\infty,x]}(X_1),\ Y_2 = I_{(-\infty,x]}(X_2),\ \ldots,\ Y_n = I_{(-\infty,x]}(X_n),$$

then

$$p = P\{Y_1 = 1\} = P\{X_1 \le x\} = F_X(x).$$

Let $S_n = Y_1 + Y_2 + \cdots + Y_n$, then $S_n \sim Bin(n,p)$. Also, $S_n \ge j$ if and only if at least j of the $X_i \le x$. In turn, this is true if and only if the j-th order statistic $X_{(j)} \le x$. Consequently,

$$\begin{aligned} F_{X(j)}(x) &= P\{X_{(j)} \le x\} = P\{S_n \ge j\} \\ &= \sum_{k=j}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=j}^{n} \binom{n}{k} F_X(x)^k (1 - F_X(x))^{n-k} \end{aligned}$$

Marginal Distribution Functions

For example, the maximum of the X_i has distribution function $F_{X_{(n)}}(x) = F_X(x)^n$. If the X_i are $U(0,1)$, the distribution of the maximum is x^n on the interval $[0,1]$.

The minimum $X_{(1)}$ has distribution function

$$F_{X_{(1)}}(x) = P\{X_{(1)} \leq x\} = 1 - P\{X_{(1)} > x\} = 1 - \prod_{i=1}^{n} P\{X_i > x\} = 1 - (1 - F_X(x))^n$$

. If the X_i are $Exp(\beta)$, the distribution is the minimum is

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n = 1 - \exp(-\beta x)^n = 1 - \exp(-n\beta x)$$

and the minimum is $Exp(n\beta)$.

We can find the density of the order statistics by differentiation

$$f_{X(j)}(x) = \frac{d}{dx} F_{X(j)}(x).$$

However, we will take a more intuitive approach.

Distribution Functions

For

$$f_{X_{(j)}}(x)\Delta x \approx P\{x \le X_{(j)} < x + \Delta x\}$$

to hold, we must have

- one of the X_i in the interval from x to $x + \Delta x$,
  n mutually exclusive choices each with probability $f_X(x)\Delta x$,
- $j - 1$ of the X_i in the interval up to x,
  $\binom{n-1}{j-1}$ mutually exclusive choices each with probability $F_X(x)^{j-1}$
- $n - j$ of the X_i in the interval above to $x + \Delta x$.
  1 choice with probability $(1 - F_X(x + \Delta x))^{n-j}$

Thus,

$$f_{X_{(j)}}(x)\Delta x \approx n f_X(x)\Delta x \binom{n-1}{j-1} F_X(x)^{j-1}(1 - F_X(x + \Delta x))^{n-j}$$

$$f_{X_{(j)}}(x) = \binom{n}{j} j F_X(x)^{j-1}(1 - F_X(x)^{n-j} f(x).$$

Distribution Functions

Exercise. $n\binom{n-1}{j-1} = \binom{n}{j}j$.

$$n\binom{n-1}{j-1} = n\frac{(n-1)_{j-1}}{(j-1)!} = \frac{(n)_j}{j!}j = \binom{n}{j}j.$$

If the U_i are uniform random variables on the interval $[0,1]$, then

$$\begin{aligned} f_{U_{(j)}}(u) &= \binom{n}{j} j u^{j-1}(1-u)^{n-j}, \qquad 0 \le u \le 1 \\ &= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} u^{j-1}(1-u)^{n-j}, \end{aligned}$$

a beta random variable with parameters $\alpha = j$ and $\beta = n-j+1$
Thus,

$$EU_{(j)} = \frac{j}{n+1} \quad \text{and} \quad \mathrm{Var}(U_{(j)}) = \frac{j(n-j+1)}{(n+1)^2(n+2)}.$$

Bivariate Distribution Functions

Let repeat the ideas for one dimensional marginal distributions to obtain the distribution for two dimensional marginals. To begin, pick $1 \leq i < j \leq n$ and $x_i < x_j$
For

$$f_{X_{(i)},X_{(j)}}(x_i, x_j)(\Delta x)^2 \approx P\{x_i \leq X_{(i)} < x_i + \Delta x, x_j \leq X_{(j)} < x_j + \Delta x\}$$

to hold, we must have

- one X_k in the interval $(x_i, x_i + \Delta x)$.
  n mutually exclusive choices each with probability $f_X(x_i)\Delta x$,
- and another in the interval from $(x_j, x_j + \Delta x)$.
  $n - 1$ mutually exclusive choices each with probability $f_X(x_j)\Delta x$,
- $i - 1$ of the X_k in the interval up to x_i,
  $\binom{n-2}{i-1}$ mutually exclusive choices each with probability $F_X(x_i)^{i-1}$

Thus,

$$f_{X_{(j)},X_{(j)}}(x_i, x_j)(\Delta x)^2 \approx nf_X(x_i)\Delta x(n-1)f_X(x_j)\Delta x \binom{n-2}{i-1} F_X(x_i)^{i-1}$$

Bivariate Distribution Functions

- $j-i-1$ of the X_k in the interval between $x_i+\Delta x$ and x_j
  $\binom{n-i-1}{j-i-1}$ mutually exclusive choices each with probability
  $(F_X(x_j)-F_X(x_i+\Delta x))^{j-i-1}$
- $n-j$ of the X_k in the interval above to $x_j+\Delta x$.
  1 choice with probability $(1-F_X(x_j+\Delta x))^{n-j}$

Thus, $f_{X_{(i)},X_{(j)}}(x_i,x_j)(\Delta x)^2$

$$
\begin{aligned}
\approx\ & (n)_2 f_X(x_i) f_X(x_j)(\Delta x)^2 \binom{n-2}{i-1} F_X(x_i)^{i-1} \\
& \times \binom{n-i-1}{j-i-1}(F_X(x_j)-F_X(x_i+\Delta x))^{j-i-1}(1-F_X(x_j+\Delta x))^{n-j}
\end{aligned}
$$

$$
f_{X_{(i)},X_{(j)}}(x_i,x_j) = (n)_2 \binom{n-2}{i-1,\, j-i-1,\, n-j} f_X(x_i) f_X(x_j) F_X(x_i)^{i-1} (F_X(x_j)-F_X(x_i))^{j-i-1} (1-F_X(x_j))^{n-j}
$$

Bivariate Distribution Functions

$$f_{X_{(j)},X_{(j)}}(x_i, x_j) = (n)_2 \binom{n-2}{i-1, j-i-1, n-j} f_X(x_i) f_X(x_j) F_X(x_i)^{i-1} (F_X(x_j) - F_X(x_i))^{j-i-1} (1 - F_X(x_j)$$

For $U(0,1)$ variables,

$$\begin{aligned} f_{U_{(j)},U_{(j)}}(u_i, u_j) &= (n)_2 \binom{n-2}{i-1, j-i-1, n-j} u_i^{i-1} (u_j - u_i)^{j-i-1} (1-u_j)^{n-j} \\ f_{X_1,X_2}(x_1, x_2) &= \frac{\Gamma\left(\sum_{i=1}^{3} \alpha_i\right)}{\prod_{i=1}^{3} \Gamma(\alpha_i)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} (1 - x_1 - x_2)^{\alpha_3 - 1} \end{aligned}$$

where $x_1 = u_i, x_2 = u_j - u_i$, $\alpha_1 = i, \alpha_2 = j - i, \alpha_3 = n - j + 1$.

Dirichlet Random Variables

The random vector $X = (X_1 \ldots, X_m)$ is called a Dirichlet random variable ($X \sim Dir(\alpha)$) with parameter $\alpha = (\alpha_1, \ldots, \alpha_m), \alpha_i > 0$ provided that its density

$$f_X(\mathbf{x}|\alpha) = \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} x_1^{\alpha_1 - 1} \ldots x_m^{\alpha_m - 1}$$

for $\mathbf{x} = (x_1, \cdots, x_m)$ on the $m-1$ simplex $\Delta^{m-1} = \{\mathbf{x} \in \mathbb{R}^m; x_i \geq 0, \sum_{i=1}^m x_i = 1\}$.

- For $\alpha = (1, \ldots, 1)$, X is uniform on Δ^{m-1}.
- $(X_1 \ldots, X_i + X_{i+1}, \ldots X_m) \sim Dir(\alpha_1, \ldots, \alpha_i + \alpha_{i+1}, \ldots, \alpha_m)$
- For $|\alpha| = \alpha_1 + \cdots + \alpha_m$,

$$X_i \sim Beta(\alpha_i, |\alpha| - \alpha_i), \quad EX_i = \frac{\alpha_i}{|\alpha|}, \quad \mathrm{Var}(X_i) = \frac{\alpha_i(|\alpha| - \alpha_i)}{|\alpha|^2(|\alpha| + 1)}$$

Thus, for large values of $|\alpha|$, the X_i are more concentrated about its mean.

Bivariate Distribution Functions

$$\text{Cov}(X_i, X_j) = -\frac{\alpha_i \alpha_j}{|\alpha|^2(|\alpha|+1)}$$

Returning to the order statistics, where $x_1 = u_i, x_2 = u_j - u_i$, $\alpha_1 = i, \alpha_2 = j - i, \alpha_3 = n - j + 1$.

$$EU_{(i)} = \frac{i}{n+1}, \quad \text{Var}(U_{(i)}) = \frac{i(n+1-i)}{(n+1)^2(n+2)}$$

$$\begin{aligned}
\text{Cov}(U_{(i)}, U_{(j)}) &= \text{Cov}(U_{(i)}, U_{(i)} + (U_{(j)} - U_{(i)})) \\
&= \text{Var}(U_{(i)}) + \text{Cov}(U_{(i)}, U_{(j)} - U_{(i)}) \\
&= \frac{i(n+1-i)}{(n+1)^2(n+2)} - \frac{i(j-i)}{(n+1)^2(n+2)} = \frac{i(n+1-j)}{(n+1)^2(n+2)}
\end{aligned}$$

Bivariate Distribution Functions

$$\mathrm{Cov}(U_{(i)}, U_{(j)}) = \frac{i(n+1-j)}{(n+1)^2(n+2)}$$

The correlation

$$\rho_{U_{(i)},U_{(j)}} = \frac{i(n+1-j)}{\sqrt{i(n+1-i)}\sqrt{j(n+1-j)}} = \sqrt{\frac{i}{n+1-i}} \cdot \sqrt{\frac{n+1-j}{j}}$$

$$= \sqrt{\frac{i/(n+1)}{1-i/(n+1)}} \Bigg/ \sqrt{\frac{j/(n+1)}{1-j/(n+1)}}, \quad i < j$$

Introduction

In the simplest possible terms, the goal of estimation theory is to answer the question:

What is that number?

Statistics has provided two distinct approaches this question - typically called

- classical or frequentist, and
- Bayesian.

Definition. A statistic is a function of the data that does not depend on any unknown parameter.

Parameter Estimation

For parameter estimation, we consider $X = (X_1, \ldots, X_n)$, independent random variables chosen according to one of a family of probabilities P_θ where θ is element from the parameter space Θ. Based on our analysis, we choose an estimator $\hat{\theta}(X)$. If the data $\mathbf{x}$ takes on the values $x_1, x_2, \ldots, x_n$, then

$$\hat{\theta}(x_1, x_2, \ldots, x_n)$$

is called the estimate of θ. Thus we have three closely related objects.

1. θ - the parameter, an element of the parameter space, is a number or a vector.
2. $\hat{\theta}(x_1, x_2, \ldots, x_n)$ - the estimate, is a number or a vector obtained by evaluating the estimator on the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
3. $\hat{\theta}(X_1, \ldots, X_n)$ - the estimator, is a random variable. We will analyze the distribution of this random variable to decide how well it performs in estimating θ.

Parameter Estimation

For Bernoulli trials $X = (X_1, \ldots, X_n)$, we have

1. p, a single parameter, the probability of success, with parameter space $[0, 1]$.
2. $\hat{p}(x_1, \ldots, x_n)$ is the sample proportion of successes in the data set.
3. $\hat{p}(X_1, \ldots, X_n)$, the sample mean of the random variables

$$\hat{p}(X_1, \ldots, X_n) = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}S_n$$

is an estimator of p. We can give the distribution of this estimator because S_n is a binomial random variable.

Parameter Estimation

In classical statistics, the state of nature is assumed to be fixed, but unknown to us. Thus, one goal of estimation is to determine which of the P_θ is the source of the data. The estimate is a statistic

$$\hat{\theta} : \text{data} \to \Theta.$$

For estimation procedures, the classical approach to statistics is based on two fundamental questions:

- How do we determine estimators?
- How do we evaluate estimators?
  - Does this estimator in any way systematically under or over estimate the parameter?
  - Does it has large or small variance?
  - How does it compare to a notion of best possible estimator?
  - How easy is it to determine and to compute?
  - How does the procedure improve with increased sample size?

Estimating Means

Beginning with independent samples $x_1, \ldots, x_n$ from an unknown distribution $F_X(x)$, then, sample mean $\bar{x} = (x_1 + \cdots + x_n)/n$ the most obvious statistic to estimate for the distributional mean μ.

For $X_1, \ldots, X_n$ independent with distribution $F_X(x)$, we have learned

- $E\bar{X} = \mu$
- $\mathrm{Var}(\bar{X}) = \sigma^2/n$

This first property is a statement that $\bar{X}$ is an unbiased estimator for μ

We will consider the circumstance for estimating the variance σ^2

Estimating Variances

We begin with a numerical investigation into estimating the variance. Thus, consider $x_1, \ldots, x_{10}$, 10 sample taken from 36 tosses of a fair coin, i.e., samples from a $Bin(36, 1/2)$. This distribution has mean $\mu = 36 \cdot 1/2 = 18$ and variance $\sigma^2 = 36 \cdot 1/2 \cdot (1 - 1/2) = 9.$

```
> ssx<-rep(0,5000);ssmu<-rep(0,5000)
> for (i in 1:5000){x<-rbinom(10,36,0.5);ssx[i]<-sum((x-mean(x))^2);
   ssmu[i]<-sum((x-18)^2)}
> mean(ssx);mean(ssmu)
[1] 80.92746
[1] 89.7482
> min(ssmu-ssx)
[1] 0
```

Define $SS(\alpha) = \sum_{i=1}^{n}(x_i - \alpha)^2$, this leads to two conjectures, (1) $SS(\mu) \geq SS(\bar{x})$

Estimating Variances

and

$$(2) \qquad \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \quad \text{and} \quad \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

are unbiased estimators of σ^2.

Exercise. $\min_\alpha SS(\alpha) = SS(\bar{x})$.

$$\begin{aligned} SS(\alpha) &= \sum_{i=1}^{n}((x_i-\bar{x})+(\bar{x}-\alpha))^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + 2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\alpha) + \sum_{i=1}^{n}(\bar{x}-\alpha)^2 \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\alpha)^2 = SS(\bar{x}) + n(\bar{x}-\alpha)^2. \end{aligned}$$

Estimating Variances

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}n\sigma^2 = \sigma^2$$

Finally, using the identity above with $\alpha = \mu$

$$\begin{aligned}
E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] &= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \frac{1}{n-1}\left(E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] - nE[(\bar{X}-\mu)^2]\right) \\
&= \frac{1}{n-1}(n\sigma^2 - n(\sigma^2/n)) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
\end{aligned}$$

However, because the square root is concave. Thus, by Jensen's inequality $S = \sqrt{S^2}$ is biased downward, i.e., $ES \leq \sigma$.